REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
		FINAL – 30 NOV 92 TO 14 MAY 96

4. TITLE AND SUBTITLE
MATHEMATICAL PROCESSING FOR SIMULATION AND CONTROL OF SEMICONUCTOR PROCESSING

5. FUNDING NUMBERS
F49620-93-1-0062
9806/12   61103D

6. AUTHOR(S)

K.S. TSAKALIS AND P.E. COURCH

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
CENTER FOR SYSTEMS SCIENCE AND ENGINEERING
ARIZONA STATE UNIVERSITY
BOX 877606
TEMPE, AZ 85287-7606

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NM
110 DUNCAN AVE, SUITE B115
BOLLING AFB DC 20332-8080

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

F49620-93-1-0062

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
APPROVED FOR PBULIC RELEASE:
DISTRIBUTION UNLIMITED

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)
SEE REPORT FOR ABSTRACT

19960801 097

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

NSN 7540-01-280-5500

DTIC QUALITY INSPECTED 1

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

DISCLAIMER NOTICE

UNCLASSIFIED

Technical Report
distributed by
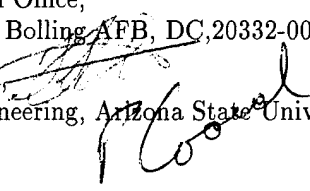
DEFENSE
TECHNICAL
INFORMATION
CENTER

DTIC

UNCLASSIFIED
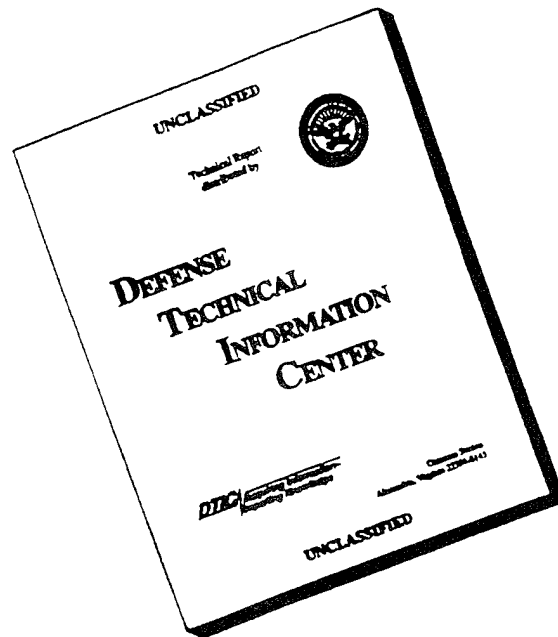
THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

Center for Systems Science and Engineering

Arizona State University

Box 877606
Tempe, AZ 85287-7606
602/965-8382

Date:	7/1/96
To:	Carmen Calvert, Contracting Officer, AFOSR/PKA 110 Duncan Ave., Suite B115, Bolling AFB, DC 20332-0001
	Dr. Marc Q. Jacobs, Technical Office, 110 Duncan Ave., Suite B115, Bolling AFB, DC,20332-001
From:	K.S. Tsakalis and P.E. Crouch Department of Electrical Engineering, Arizona State University
Subject:	Final Technical Report
Grant No.:	F49620-93-1-0062
Grant Title:	*Mathematical Processing for Simulation and Control of Semiconductor Processing*
Program Manager:	Dr. Marc Q. Jacobs, AFOSR Dr. Anna Tsao, ARPA, Defense Sciences Office
P.I.:	P.E. Crouch
Due Date:	7/14/96

Contents

1 Personnel

Since its award, seven faculty members and twelve students from the departments of Electrical Engineering, Chemical Engineering and Mathematics were funded by this project (including one post-doctoral student and three students funded by ASSERT awards related to this grant).

More specifically, the faculty members involved in this project were:

1. T. Cale, Chemical Engineering

2. P. Crouch, Electrical Engineering

3. D. Ferry, Electrical Engineering

4. M. Kozicki, Electrical Engineering

5. G. Raupp, Chemical Engineering

6. C. Ringhofer, Mathematics

7. K. Tsakalis, Electrical Engineering

Students that were supported by the grant (at various stages) and their current status is listed below:

1. R. Bammi, Ch.E., (T. Cale), Left after completing his M.S. degree.

2. S.-J. Yang, Ch.E., (T. Cale), Left project.

3. M. Peters, Ch.E., (T. Cale), Left project.

4. Y.-L. Lin, Ch.E., (G. Raupp, K. Tsakalis, P. Crouch), Left after completing his M.S. degree, currently with Texas Instruments, Dallas.

5. M. Gobbert, Math. (C. Ringhofer), Completed his PhD, currently with IMA, Univ. of Minnesota.

6. S. Shen, E.E., (P. Crouch), Left after obtaining employment with Texas Instruments, Dallas.

7. L. Song, E.E. (K. Tsakalis), Continuing towards her Ph.D. degree (expected graduation during the 1996-1997 academic year)

8. P. Thanikasalam, E.E. (D. Ferry), Continuing towards his Ph.D. degree.

The following students were supported by ASSERT awards related to the grant.

9. K. Stoddard, E.E., (P. Crouch, K. Tsakalis, M. Kozicki), Left after obtaining M.S. degree, currently employed at SEMY Eng. Inc., Phoenix.

10. K. Tracy, Ch.E., (T. Cale) Completed M.S. degree.

11. Delbert Herald, E.E. (M. Kozicki, K. Tsakalis), replaced K. Stoddard and currently continuing towards his M.S. degree (expected graduation during the 1996-1997 academic year)

Finally, J.H. Park (Ch.E.) worked as a post-doctoral student with T. Cale on the development of simulation test-beds.

2 Summary of Results

As indicated by its title, the project had three main thrust areas, namely modeling, simulation, and control of processes that are encountered in Semiconductor manufacturing applications. The key common theme of these areas was the use of detailed, first-principles, physical models to predict the process behavior under various conditions. Such an approach is currently considered as an attractive alternative for empirical models whose validity is limited by the range of testing conditions.

A clear and impressive demonstration of this concept can be given based on the results of this project. In particular, in [24–26], a detailed feature-scale model/simulator (EVOLVE, [3]) was used as a basis to derive a simplified model and compute a near-optimal temperature protocol for a CVD process in trenches, that minimizes processing time subject to step-coverage constraints. This protocol (temperature trajectory) was validated by simulation against the detailed test-bed (EVOLVE). The basic ingredient of this protocol, computed by employing optimal control principles, was that processing time can be reduced while maintaining a given step coverage, by performing the initial stage of the process at high temperatures and reduce the temperature as the deposition approaches closure. A simpler version of the same idea was then tested experimentally, [15–17], where an excellent agreement was found between simulated and actual behavior of the deposition process. The improvement of such a time-varying temperature protocol, compared to the customary constant processing conditions, is illustrated in Fig. 1. While further details can be found in the related references, these results demonstrate that significant process improvement can be obtained by using non-standard processing protocols, computed by means of optimization/optimal control techniques and detailed, physically-based simulation models. However, the validity and confidence in such protocols relies heavily on the validity of the models and simulators for non-standard processing conditions. Loosely speaking, it is expected that process optimization will, in general, involve the use of "extreme" processing conditions for which experimental data are unlikely to be available a priori. This observation imposes serious limitations of the applicability of purely empirical models to such a process optimization. On the other hand, detailed first-principles models possess the ability to accurately predict the process behavior, even under these extreme conditions. Thus, the usefulness of detailed, physically-based models is not limited to an abstract "improvement of our process understanding" but it has a tangible and important role in process optimization.

Returning to the project summary, our results can be classified into the three main categories of research thrusts. This distinction, however, is not always rigid since several of the results contain components that combine modeling, simulation and control concepts.

Beginning with the modeling problem, a part of our efforts was devoted to the dry oxidation of silicon and a unified description of thin and thick film oxide growth rates and interfacial structure. The proposed model invokes dissociative chemisorption in silicon at the interface between the silicon dioxide film and the substrate. This model supports the diffusing species in such a process to be molecular rather than atomic oxygen and predicts a self-limiting oxide film thickness of 0.5-0.6 nm and an inherent interfacial roughness of approximately one atomic diameter (0.3nm). Kinetic rate equations developed with this model were found to provide an excellent fit of experimental data. Further details are given in publications/reports [1,2].
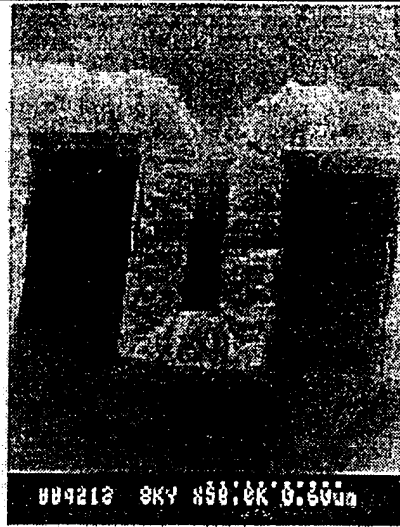
Figure 5.52 PRCVD Trench sample 5-1.

Figure 5.53 CRCVD Trench sample 6-3.

Figure 1: Experimental comparison of blanket tungsted deposition profiles under Programmed-rate and Constant-rate CVD [17]. For comparable deposition thickness and step coverage, the CRCVD processing time was 465 sec while PRCVD required only 90 sec. (80% time-savings).

Next, regarding the simulation component, our efforts were focused on the development of hybrid models, capable of describing the chemical and physical process in a single wafer reactor, on the reactor scale and the feature scale simultaneously. For this task, the challenge is the great disparity of spatial scales between reactor-scale simulators (CFDSWR, 10 cm, [4,5]) and feature-scale simulators (EVOLVE, 1 μm, [3]). To address this problem, we have introduced a "mesoscopic-scale model" on the scale ranging from several feature clusters (1 mm) to one computer chip (1 cm). This model is capable of capturing feature-to-feature and cluster-to-cluster effects that neither one of the classical reactor-scale and feature-scale model is capable of representing due to their typical length scales. The derivation of this mesoscopic-scale model relies on an asymptotic analysis of the boundary layer and utilizes homogenization techniques. This analysis yields a mathematically rigorous derivation of the model equations that provide a computationally manageable approach to self-consistently simulate the interaction of the gas flow on the reactor-scale and the global evolution of the wafer surface. Further details are given in publications/reports [6–11], containing results for the two- and three-dimensional cases and demonstrations illustrating the effects of cluster density on the deposition process (SiO_2 from TEOS) and the effect of varying operating conditions on micro-loading.

Finally, in our work related to the control of semiconductor processes, we primarily focused on the so-called "outer-loop" control. In the outer-loop, the controller is responsible for generating high-level commands in the form of trajectories to be followed by certain on-line measurable critical process variables, e.g., temperature, pressure, flowrates. Process optimization is performed at this level by computing trajectories which yield an optimal value for a desired criterion, e.g., minimize processing time and/or material consumption, subject to process quality constraints (step coverage, uniformity) and constraints imposed by the trajectory-following ability of the local inner loops (real-time controllers, e.g., PIDs). In its simplest form, the trajectories are approximated by constants and, treating the map from the inner-loop set-points to the process outputs as a static nonlinearity, the outer loop becomes the so-called "run-to-run" controller. In our work to date, we considered applications of run-to-run control in both simulation [21–23] and experimental [19,20] environments. A key component of our approach is the use of empirical but physically-motivated models to describe the process accurately (for control purposes), using relatively few adjustable parameters [22,23]. In this way, the model parameters can be adjusted using a relatively small number of data points obtained from experiments or simulation of the detailed physically-based models.

A significant portion of our efforts was devoted to the outer-loop control of LPCVD processes using optimal control principles. We considered the processes of blanket tungsten deposition and thermally activated deposition of SiO_2 from TEOS, where we derived a "control oriented," feature-scale model based on power series expansions. This model was shown to be adequate for the computation of near-optimal temperature trajectories that minimized processing time subject to step coverage constraints [?, ?]. The process improvement with the resulting processing protocols was verified with simulations against an accurate feature-scale simulation test-bed (EVOLVE, [3]) coupled with a reactor-scale simulator (CFDSWR, [4]). Preliminary experimental results in tungsten deposition have shown a similar process improvement with these protocols, relative to constant processing conditions [15–18].

On the other hand, detailed simulation test-beds of the processes and associated equipment, need to be developed simultaneously with the control oriented models. Such a concurrent development aims not only to assess the performance of the process step during design or modification, but also to provide guidance in the development of control algorithms. For example, a detailed physical model may suggest the explicit introduction of certain terms in the control oriented model, that captures the main nonlinearities or dynamics in the process behavior [22,23]. This idea was used in [27] to effectively translate the computed optimal processing conditions at the wafer surface (temperature and partial pressures) into reactor-scale manipulated variables (temperature and flow rates). A rigorous experimental verification of the process improvement with these control algorithms is currently under investigation [18].

In conclusion, our work on this project has addressed most of the original objectives and objectives that became important in the course of our research. Our results demonstrated the practical feasibility and benefits of a systematic and rigorous application of numerical analysis and control theoretic principles to the modeling, simulation and control of semiconductor manufacturing processes. A by-product, of perhaps equal importance, is the stimulation of a closer collaboration of the PI's with the local semiconductor industry that has resulted in several small application-oriented projects, sponsored by, e.g., Intel, Motorola, Semy.

3   Publications and Reports

1. T. K. Whidden, P. Thanikanasalam, M. J. Rack, and D. K. Ferry, "The Initial Oxidation of Silicon(100): A Unified, Chemical Model for Thin and Thick Oxide Growth Rates and Interfacial Structure," *22nd Conference on the Physics and Chemistry of Semiconductor Interfaces*, Scottsdale, AZ, January 1995; also in *J. Vac. Sci. Technol. B*, 13, 4, 1618–1625, Jul/Aug 1995.

2. P. Thanikanasalam and D. K. Ferry, "A Unified Chemical Model for Thermal Oxidation of Silicon (100) in a Dry Oxygen Ambient," *Technical Report*, May 1996.

3. T. S. Cale, *EVOLVE 4.0b, A low pressure deposition simulator.* August 1994.

4. J.H. Park, *User's guide for CFDSWR: Computational Fluid Dynamics of a Single Wafer Reactor.* Arizona State Univ. 1993.

5. M. Gobbert: "The Simulation Platform CMFSD," Report, *Arizona State University*, 1993.

6. C. Ringhofer, "A New Expansion Procedure to Generalize Hydrodynamic Transport Models," *Proc. International Workshop on Computational Electronics*, Portland OR, May 1994.

7. M. Gobbert, C. Ringhofer, "An Asymptotic Analysis for a model of Chemical Vapor Deposition on a micro Structured Surface," Submitted to *SIAM Journal on Applied Mathematics*.

8. M. Gobbert, T. Cale, C. Ringhofer, "One Approach to Combining Equipment Scale and Feature Scale Models," *Proc. 187th. meeting of the Electro. - Chemical Society*, Reno, Nevada, March 1995.

9. M. K. Gobbert, T. S. Cale and C. Ringhofer, "The Combination of Equipment Scale and Feature Scale Models for Chemical Vapor Deposition via a Homogenization Technique," *4th International Workshop on Computational Electronics*, Tempe, AZ, November, 1995.

10. M.K. Gobbert and C. Ringhofer, "Mesoscopic Scale Modeling of Microloading During LPCVD," *J. Electrochem. Soc.*, Aug. 1996 (to appear).

11. M. Gobbert, *A Homogenization Technique for the Development of Mesoscopic Scale Models for Chemical Vapor Deposition*. Ph.D. Dissertation, ASU, April 1996.

12. R. Bammi, T. S. Cale and G. Grivna, "Development of a Gate Metal Etch Process for Gallium Arsenide Wafers," *Thin Solid Films*, 253, 501, 1994.

13. R. Bammi, *Development of a Gate Metal Etch Process for Gallium Arsenate*, MS Thesis, Arizona State University, Arizona, April 1994.

14. T. S. Cale and Yu-Kuang Chang, "Experimental and Simulation Study of Re-emission during Sputter Deposition of Ti-W Films," to be presented at the *42nd National Symposium and Topical Conferences of American Vacuum Society*, 1996.

15. K. M. Tracy, S. Bolnedi, T. S. Cale, "Programmed Rate Chemical Vapor Deposition of Blanket Tungsten Thin Films," to be presented at the *Proc. 12th International VLSI Multilevel Interconnection Conf.*, T. Wade Ed., VMIC, 643, 1995.

16. K. M. Tracy, S. Bolnedi, G. J. Leusink and T. S. Cale, "Blanket Tungsten Film Deposition Using Programmed Rate CVD," *Advanced Metallization and Interconnect Systems for ULSI Applications in 1995*, MRS, in press.

17. K.M. Tracy, *Programmed Rate Chemical Vapor Deposition: Blanket Tungsten Film Characterization*. M.S. Thesis, ASU, May 1996.

18. J. Kristof, K. Tracy, L. Song, K. Tsakalis and T. Cale, "Programmed Rate Chemical Vapor Deposition of Tungsten," to appear in *Tech.Con. '91, SRC*, Phoenix, Sept. 1996.

19. K. Stoddard, P. Crouch, M. Kozicki, and K. Tsakalis, "Application of Feedforward and Adaptive Feedback Control to Semiconductor Device Manufacturing," *Proc. IEEE American Control Conference*, 892–896, Baltimore, MD., 1994.

20. K. Stoddard. *Application of Feed-Forward and Adaptive Feedback Control to Semiconductor Device Processing*. MS Thesis, Arizona State University, Arizona, December 1994.

21. K. S. Tsakalis and L. Song, "Set-Membership Estimation for Weakly Nonlinear Models: An Application to the Adaptive Control of Semiconductor Manufacturing Processes," *Proc. IEEE Conference on Decision and Control*, 1066–1071, Orlando Florida, Dec. 1994.

22. T. Cale, P. E. Crouch, S. Shen, and K. S. Tsakalis, "A Simple Adaptive Optimization Algorithm for the Tungsten LPCVD Process," *Proc. IEEE, American Control Conference*, 1294–1298, Seattle, 1995.

23. T. S. Cale, P. E. Crouch, S. Shen and K. S. Tsakalis, "Run to Run Adaptive Optimization of a Tungsten Silicide LPCVD Process," *34th IEEE Conference on Decision and Control*, 2474–2475, New Orleans, Dec. 1995.

24. T. Cale, P. E. Crouch, L. Song, and K. S. Tsakalis, "Increasing Throughput in Low Pressure Chemical Vapor Deposition: An Optimal Control Approach," *Proc. IEEE, American Control Conference*, 1289–1293, Seattle, 1995.

25. T. Cale, P. E. Crouch, L. Song, and K. S. Tsakalis, "Optimal control for LPCVD," *Proc. E. Chem.Soc., 1st Symposium on Process Control Diagnostics & Modeling in Semiconductor Manufacturing*, 97–108, Reno, 1995.

26. P. E. Crouch, L. Song, K. S. Tsakalis and T. S. Cale, "Optimal Control Process to Increase Single Wafer Reactor Throughput in LPCVD," in *Proc. 1995, International Symposium on Semiconductor Manufacturing*, 233-238, Austin 1995.

27. L. Song, S. Shen, K. S. Tsakalis, P. E. Crouch and T. S. Cale, "Optimal Control for Increasing Throughput in Low Pressure Chemical Vapor Deposition," submitted to *J. Electrochem. Soc.*, June 1996.

28. K. S. Tsakalis, "Bursting Scenaria and Performance Limitations of Adaptive Algorithms in the Absence of Excitation," *Proc. 3rd IEEE Mediterranean Symposium on New Directions in Control and Automation*, 257–264, Limassol, Cyprus, July 1995.

29. K.S. Tsakalis "Control and Identification of Linear Time-Varying Plants Using I/O Methods," *Proc. 34th Conf. Decis. Contr.*, 2527–2532, New Orleans, Dec. 1995.

30. P. E. Crouch, A. Laib, F. Lamnabhi-Lagarrigue, "On Adaptive Feedback Stabilization for Nonlinear Systems Modeled by Discrete Time Equivalents," *NOLCOS*, Tahoe, Nevada, 1995.

4 Appendix

The Appendix contains copies of selected recent publications and reports that provide a detailed description of the results mentioned in the Summary. The publications included here are:

$$[1,2,9,10,11,15,17,18,19,21,22,23,24,25,26,27]$$

Copies of the rest of the publications that were supported by the grant are available upon request.

Initial oxidation of silicon (100): A unified chemical model for thin and thick oxide growth rates and interfacial structure

T. K. Whidden, P. Thanikasalam, M. J. Rack, and D. K. Ferry
Nanostructures Research Group, Center for Solid State Electronics Research, Arizona State University, Tempe, Arizona 85287-5706

A model for silicon oxidation that invokes dissociative chemisorption of molecular oxygen at the interface between silicon dioxide and silicon is described. The model accounts for a self-limiting oxide film thickness of 0.5–0.6 nm (for oxidations performed at temperatures sufficient to dissociate surface dimers and permit oxygen penetration of the substrate beyond a single monolayer of suboxide). Detailed examination of the model suggests a mechanism for an inherent oxide/silicon interface roughness of approximately one atomic diameter. Kinetic rate equations developed from the model successfully account for the observed power law dependence of rate on oxygen partial pressure. These relationships were used in the derivation of an expression for the variation of oxide film growth rate with overlying oxide thickness. The relationship is tested against experimental observations reported in the literature and found to give an excellent fit. © *1995 American Vacuum Society.*

I. INTRODUCTION

The oxide of silicon is a uniquely critical material within silicon semiconductor technology, in that it is a major component for metal–oxide–semiconductor (MOS) devices, as well as having importance in other metal–insulator–semiconductor (MIS) and tunneling structures.[1] Until device fabrication began to delve into the submicron regime, the understanding of oxide formation mechanisms and the accompanying parametric processing relationships that had been developed in the mid-1960s was adequate for determining the process and materials properties concerns of device performance. As device geometries have shrunk, however, a number of previously unimportant or unobserved issues have been found to require resolution in order for deep submicron and nanoscale devices to be constructed.[2]

Models employed in discussing the thermal oxidation of silicon have usually been developed within the framework originally formulated by Deal and Grove in 1965.[3] They proposed that the oxidation rate was determined by a combination of two processes. The first involved the actual chemical reaction of oxygen with silicon at the oxide/substrate interface, while the second was the diffusion of oxygen through the previously formed oxide film. The combination of these processes resulted in the formulation of the classical "linear-parabolic" rate law of Deal and Grove:

$$dx/dt = F/N_1 = (kC^*/N_1)/(1+k/h+kx/D_{\text{eff}}), \tag{1}$$

where x is the oxide thickness, F the total flux of oxidant molecules through the oxide, k the first-order rate constant for the oxidation, C^* the concentration of oxidant at the oxide surface, N_1 the number of oxidant molecules incorporated into a unit volume of the oxide layer, h the gas phase transport coefficient of oxygen, and D_{eff} the effective diffusion coefficient of oxygen in silicon dioxide. When the differential equation is solved, Eq. (1) may be rewritten as

$$x^2 + Ax = B(t+\tau), \tag{2}$$

where

$$A = 2D_{\text{eff}}(1/k+1/h), \tag{3}$$

$$B = 2D_{\text{eff}}C^*/N_1, \tag{4}$$

and

$$\tau = (x_0^2 + Ax_0)/B. \tag{5}$$

The Deal–Grove model requires the parameter τ (a shift in the time coordinate) in order to account for the presence of an initial oxide layer on the silicon surface.

The above approach was found to provide an excellent fit to experimental data for oxidation processes utilizing H_2O/O_2 mixtures and for dry oxidation processes at thicknesses in excess of ca. 40 nm.[4] Deal–Grove estimates fail, however, in predicting oxidation rates in dry oxygen within the thin regime (<20 nm) and in accounting for the observed pressure dependence of both thin and thick oxidations. Within the thin regime, observed oxidation rates in dry processes are significantly greater than Deal–Grove predictions and the reaction exhibits a power law dependence on the oxygen pressure with rate proportional to P^m where $m = 0.6–0.8$.[5,6]

A variety of models have been proposed which address the failures of the Deal–Grove model in the thin oxide limit. Space-charge effects have been postulated to give enhanced rates in thin oxidations.[7] The phenomenon requires the oxidant to exist and diffuse in *ionic* form and is operative only over oxide thicknesses comparable to the extent of the space-charge region. Based on a detailed analysis of experimental parameters, oxidation results, and the effects of the electric field at the interface on oxygen transport, it generally has been concluded that space-charge effects are not the mechanism giving rise to enhanced oxidation rates within the thin oxidation regime.[8] Other attempts at explaining elevated rates in thin oxidations proposed structural differences between thin and thick oxides, which lead to enhanced oxygen diffusion in the thin region.[9] The presence of microchannels
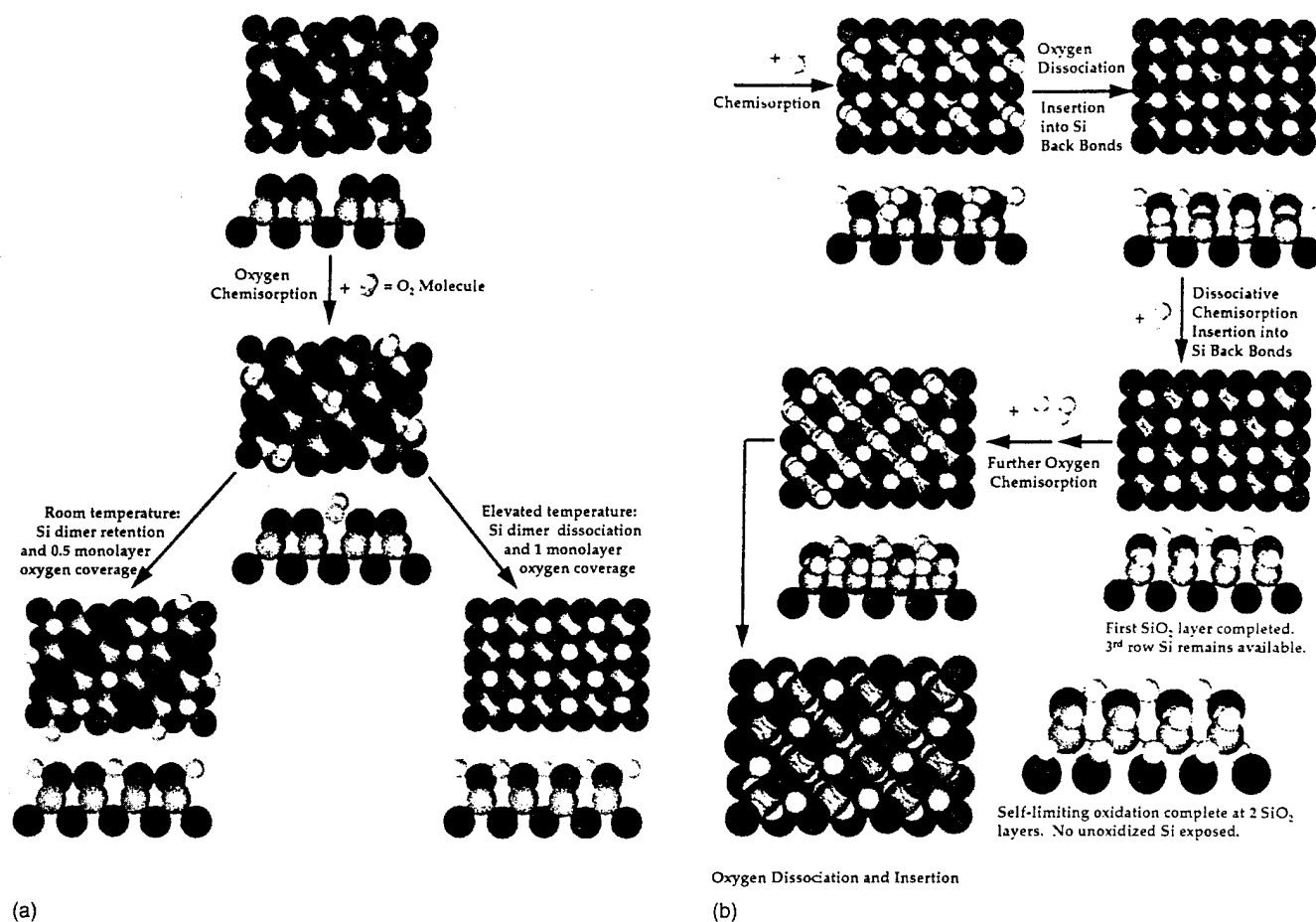
FIG. 1. (a) A conceptual model for the layer-by-layer oxidation of silicon by dry oxygen. Relative sphere sizes are proportional to elemental covalent radii, with the smaller spheres representing the oxygen atoms and the larger silicon: the initial monolayer and submonolayer oxidations. (b) A conceptual model for the layer-by-layer oxidation of silicon by dry oxygen: self-limiting oxidation of silicon at higher temperature oxidations.

with diameters of about 5 nm has been proposed to increase oxygen transport in thin oxides.[10] Some experimental evidence for such microchannels exists.[11] Stress at the oxide/substrate interface is also thought to affect the rate of oxidation.[10] Stress has been invoked to vary either the transport of oxygen through the oxide film or the intrinsic reaction rate at the interface.[12–14] Finally, variations in the oxidizing species have been suggested as potential sources of variation in oxidation rates in the thin and thick regimes.[15] Models which invoke the presence of atomic oxygen as the primary diffusing species cannot be ruled out on the basis of current evidence.

An examination of the literature and available experimental evidence does not permit a definitive choice of model for the oxidation of silicon. Here, we begin an examination of silicon oxidation with a model based on the dissociative chemisorption of molecular oxygen at the oxide/silicon interface. A fundamental rate equation is developed from this model and this determines the inherent rate of oxidation of silicon under a given set of process conditions. This equation, with appropriate modifications for the diffusion-controlled availability of oxygen at the interface, is capable of predicting oxidation rates in both thin and thick oxide regimes.

II. PROPOSED SILICON OXIDATION MECHANISM

Our conceptual model is a process in which the oxidation of silicon involves dissociative chemisorption of molecular oxygen. Figure 1(a) shows an idealized schematic of the silicon (100) surface with the classic (2×1) reconstruction. Sphere sizes in the schematic diagram reflect the relative covalent radii of silicon and oxygen. A number of different adsorption sites on the reconstructed surface are apparent in the diagram. The nature of the individual sites and the calculated energetics for oxygen adsorption have been reported in the literature.[16,17] It is instructive to approach the chemisorption of the oxygen molecule on this surface from the point of view of the molecular orbitals of the oxygen molecule. The molecular orbital energy diagram for ground state oxygen is shown in Fig. 2(a). The highest-energy, occupied molecular orbitals consist of two half-filled orthogonal and degenerate orbitals of antibonding symmetry. 95% probability surfaces for the orbitals are shown in Fig. 2(b). Interactions between the oxygen molecule and the silicon surface are initiated by the mixing of the orbitals with the highest-energy orbitals of the surface. The highest-energy orbital on the silicon (2×1) surface is postulated to be a half-filled "dangling bond" on each silicon atom in the dimer units.
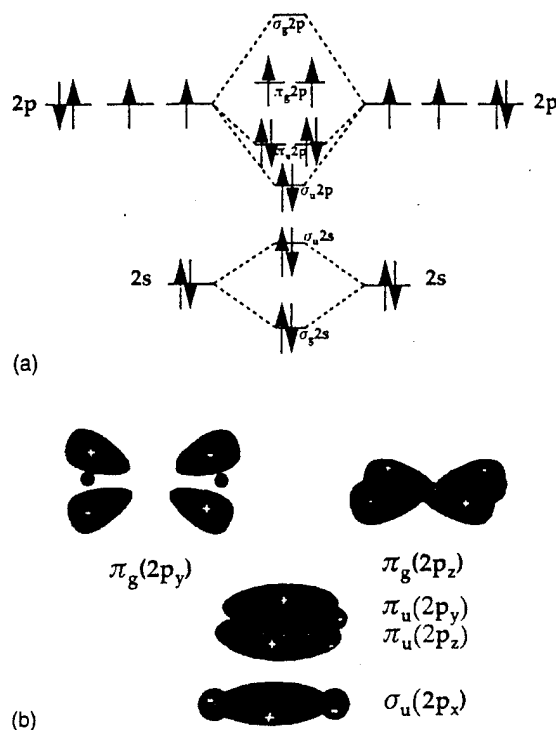
$\sigma_g 2p$

2p $\pi_g 2p$ 2p

$\pi_g 2p$

$\sigma_u 2p$

$\sigma_u 2s$

2s $\sigma_s 2s$ 2s

(a)

$\pi_g(2p_y)$ $\pi_g(2p_z)$

$\pi_u(2p_y)$
$\pi_u(2p_z)$

$\sigma_u(2p_x)$

(b)

FIG. 2. Molecular orbital (a) energy diagram for the dioxygen molecule; (b) 95% probability surfaces for molecular orbitals.

presumably of sp^3 symmetry. Simple considerations of the relative geometries of the surface "dangling bond" (expected to project from the dimer into the gap between dimer rows) and the orthogonal, half-filled π^* orbitals on the oxygen molecule suggest that maximum mixing will occur when adsorption takes place with the geometry depicted in the initial oxygen chemisorption step shown in Fig. 1(a). Since the interacting orbitals on the oxygen molecule are π^* in character, incorporation of electron density from the silicon acts to weaken the oxygen–oxygen bond and leads to dissociation of the oxygen molecule and insertion of the oxygen atoms into nearby silicon–silicon back bonds. Whether the chemisorption, bond dissociation, and bond insertion occur as a concerted reaction or in discrete steps cannot be distinguished without extensive additional experimental characterizations.

Experimental evidence from surface analytical studies conducted in our laboratories and other facilities[18] suggests that, under anhydrous conditions, three different initial reactions may occur, depending upon the temperature employed in the oxidation process. Figure 1(a) shows a possible reaction scheme for the dissociation of molecular oxygen on the silicon surface at room temperature. At low temperature, even though Si–O bond formation may be thermodynamically favored, there appears to be insufficient energy for dissociation of the silicon dimer bonds, and the reaction may be kinetically limited to the formation of oxygen bridges between dimers, as shown, with oxygen coverage of ca. 0.5 monolayer (Table I). Experimental evidence supporting the presence of such a structure has been reported.[18] As the temperature of the process is raised, sufficient energy for the

TABLE I. Auger spectral analyses for HF-treated and sputter-cleaned Si(100) surfaces.

	Ratio of Si_{SiO}/Si_{SiO_2}	Oxide coverage ML (Å)	Reacted oxygen (%)
HF treated			
60 min dry oxygen	0.07	0.14, 0.32	49.1
60 min ambient air	0.19	1.02, 2.36	84.7
Sputter cleaned			
60 min dry oxygen	0.33	1.16, 2.69	90.7
60 min ambient air	0.30	1.77, 4.10	105.7

dissociation of the surface dimers may become available and (1×1) oxygen surface superlattices can evolve from more extensive oxygen coverage (1 monolayer). Again, limited experimental evidence for the formation of such structures is available in the literature.[19] The exact temperature at which such a structure can form has not been determined, but a cursory examination of relative bond strengths permits an estimate of the temperature range for the formation of the fully oxidized surface. Si–H bonds are known to dissociate at temperatures of 450 °C and these have a dissociation energy of 80–90 kCal/mol. Si–Si bond dissociation energies for the surface dimers have not been determined, but it is instructive to compare them with the Si–Si dissociation energy in disilane, 74 kCal/mol. It is reasonable to expect that this energy represents an upper limit on the dissociation energy of the surface dimers. Consequently, silicon surface dimers may reasonably be expected to dissociate at temperatures around 300–400 °C. Experimental evidence on the formation of full oxygen monolayers in some intermediate temperature oxidations would appear to support this projection.[20]

Atmospheric pressure oxidations carried out at higher temperatures (but below temperatures at which sustained oxidation occurs) appear to self-limit at thicknesses of somewhat less than 1 nm. Typical values of 0.6–0.7 nm have been reported. A possible rationale for these observations is shown in Fig. 1(b), where a mechanism which utilizes the iterative action of molecular oxygen chemisorption at the oxide/silicon interface followed by O_2 bond dissociation and insertion into silicon back bonds produces a layer-by-layer growth of the oxide film. The dependence of the mechanism on the chemisorption of molecular oxygen via orbital mixing with the unoxidized silicon at the interface necessarily limits the extent to which the reaction can proceed without some additional driving force. It may be seen from Fig. 1(b) that unoxidized silicon becomes unavailable to surface adsorbed oxygen after the formation of two fully oxidized layers of silicon, and the reaction will therefore cease at this point for any temperature below those at which appreciable oxygen diffusion to the interface can occur. The calculated thickness of this self-limited layer of oxide films is 0.5–0.6 nm, consistent with the experimentally observed thickness in self-limited oxide films.

This model thus appears to adequately describe the structural phenomena that have been reported for dry oxidations of clean silicon surfaces. It must be noted that the model does not address oxidations in which moisture is present, nor does it address oxidations and native oxide thicknesses on surfaces that have been exposed to moisture at any time prior
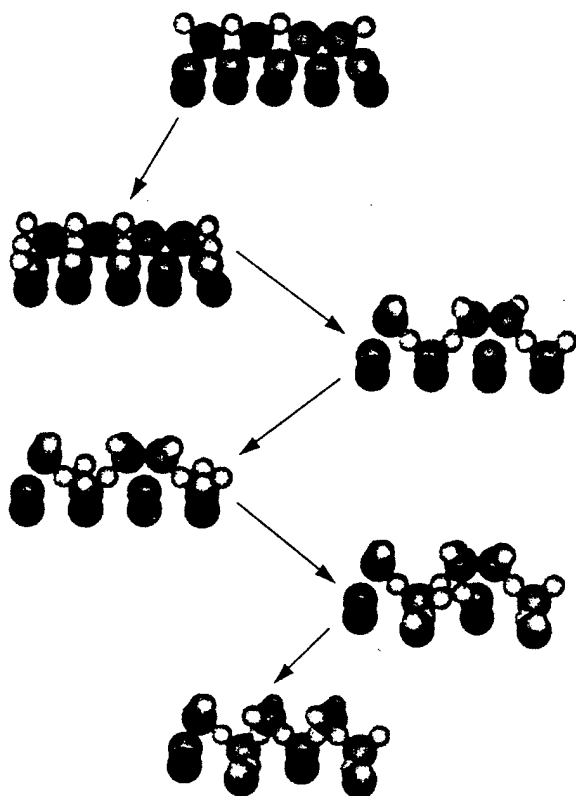
FIG. 3. A possible mechanism for the origin of inherent interfacial roughness between the silicon dioxide film and the substrate (see the text).

to dry oxidation. In such cases, distinctly different mechanisms of oxidation can be expected due to the presence of water.

The model may also provide a mechanistic rationale for the observation that oxide/silicon interfaces appear to exhibit an inherent interfacial roughness of about one atomic diameter. Consider Fig. 3, in which the (011) projection of the silicon (100) surface is shown during successive stages of oxidation. At the top of the figure a mixed configuration for the initial oxidation layer is depicted in which both the (1×1) and (2×1) oxygen bridged structures are present on the surface. Such a configuration might arise when oxidation of the silicon is permitted during transient temperature ramping. In such an arrangement, areas of the surface that have retained the (2×1) structure effectively cause a temporary barrier to oxygen insertion in the underlying back bonds, while oxidation may proceed in the areas with the more open (1×1) oxygen/silicon arrangement. Oxidation of the top layer back bonds gives a juncture of the two types of surface structure as shown in the second and third schematics of the figure. As the oxidation proceeds, and the dimer is eventually oxidized, the discrepancy between the oxidation depth under initially (1×1) and (2×1) surfaces may be expected to propagate with the interface, giving rise to an inherent interfacial roughness of 0.3 nm, approximately the diameter of an SiO grouping. Oxidation to the final condition exhibiting such interfacial roughening is depicted in the lower three schematics of Fig. 3.

III. SILICON OXIDATION KINETICS

Utilizing the conceptual model developed in the previous discussion, rate equations for the oxidation of silicon by dry oxygen can be developed. The model suggests that the reaction occurs via classical Langmuir kinetics, with the production of an intermediate species prior to the final reaction to product SiO_2. We currently formulate this intermediate as "SiO," although it is acknowledged that a completely rigorous treatment would require a more extensive analysis of the intermediate species and a subsequently more detailed series of equations to describe the chemisorption process. Such rigorous treatments will be pursued in future work. Following the Langmuir formalism, the reaction sequence for silicon oxidation may be simply written:

$$Si + \tfrac{1}{2}O_2 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} SiO,$$

$$SiO + \tfrac{1}{2}O_2 \overset{k_2}{\rightarrow} SiO_2,$$
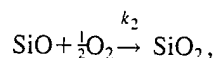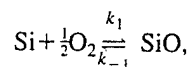
where the first reaction represents the equilibrium between molecular oxygen chemisorbed at the substrate/oxide interface and the substrate silicon and the second is an irreversible reaction of the intermediate with additional chemisorbed oxygen to yield the product SiO_2. k_1, k_{-1}, and k_2 are the usual rate constants. From the second of these equations, the rate of formation of the product, $d[SiO_2]/dt$ may be expressed as

$$d[SiO_2]/dt = k_2[SiO][O_2]^{1/2}, \tag{6}$$

where [SiO] is the concentration of the partially oxidized intermediate at the interface and $[O_2]$ the concentration of chemisorbed molecular oxygen available at the interface. The concentration of oxygen can be determined using diffusional considerations similar to those employed in the Deal–Grove model and most other current reaction models. Steady-state assumptions are not valid in this reaction since the availability of oxygen in the equilibrium reaction will vary within the reaction time frame due to diffusion limitations. The usual equilibrium relationships therefore will not be considered. However, absolute expressions for reaction rates remain valid and the concentration of the intermediate species may be determined by examining its rate of change:

$$d[SiO]/dt = k_1[Si][O_2]^{1/2} - k_{-1}[SiO] - k_2[SiO][O_2]^{1/2}. \tag{7}$$

We know that the concentration of the SiO intermediate at the interface is a constant after the formation of a thick layer of oxidized silicon; therefore $d[SiO]/dt = 0$ and so, in this limit,

$$[SiO] = \frac{k_1[Si][O_2]^{1/2}}{k_{-1} + k_2[O_2]^{1/2}}. \tag{8}$$

Substituting this expression into Eq. (1), the rate of formation of the silicon dioxide film can be expressed as

$$d[SiO_2]/dt = \frac{k_1 k_2[Si][O_2]}{k_{-1} + k_2[O_2]^{1/2}}. \tag{9}$$

and the growth rate of the film may be written as

$$Ndx/dt = \frac{k_1 k_2 [\text{Si}][\text{O}_2]}{k_{-1} + k_2 [\text{O}_2]^{1/2}}, \tag{10}$$

where N is a conversion factor for film thickness to $[\text{SiO}_2]$ analogous to that employed in the Deal–Grove model. Since $[\text{Si}]$ is a constant in the system, it may be seen that the film growth rate is controlled solely by the oxygen concentration at the oxide/substrate interface. Equation (10) results from the fundamental reaction-rate equation (9) and is applicable for the oxidation of silicon by dry oxygen in the thick limit. It is precisely this function that must be modified for thin oxides in the transient regime.

To allow for the non-steady-state behavior of the oxygen concentration, we return to Eq. (7) and multiply both sides of the equation by the integrating factor $e^{\gamma t}$ and, letting $k_{-1} + k_2 [\text{O}_2]^{1/2} \equiv \gamma$, we can express the derivative of the intermediate concentration with respect to time as

$$d/dt\{[\text{SiO}]e^{\gamma t}\} = \{k_1 [\text{Si}][\text{O}_2]^{1/2}\} e^{\gamma t}. \tag{11}$$

Integration gives

$$[\text{SiO}] = \frac{k_1 [\text{Si}][\text{O}_2]^{1/2}}{k_{-1} + k_2 [\text{O}_2]^{1/2}} (1 - e^{-\gamma t}). \tag{12}$$

Substituting into Eq. (1), the rate of formation for silicon dioxide can then be expressed as:

$$d[\text{SiO}_2]/dt = Ndx/dt = \frac{k_1 k_2 [\text{Si}][\text{O}_2]}{k_{-1} + k_2 [\text{O}_2]^{1/2}} (1 - e^{-\gamma t}). \tag{13}$$

We note that this latter equation predicts no $[\text{SiO}]$ for thin oxides (small t). We return to this point below. Setting $\gamma t = x/a$, which equates the total amount of intermediate reacted to produce the oxide film with the film thickness divided by some unit length a, the relationship for film thickness versus time becomes

$$Ndx/dt = \frac{k_1 k_2 [\text{Si}][\text{O}_2]}{k_{-1} + k_2 [\text{O}_2]^{1/2}} (1 - e^{-x/a}). \tag{14}$$

Relation (14) describes the growth of the oxide in terms of the concentration of molecular oxygen at the oxide/substrate interface. Since this concentration cannot be experimentally determined, we need to express the interfacial concentration of oxygen in terms of the oxygen concentration at the oxide surface, a known quantity that may be calculated from the oxygen partial pressure in the gas phase above the substrate. To obtain this relationship we make use of Fick's laws, and the continuity of flux, in a derivation analogous to that of the Deal–Grove model. The three fluxes are expressed, in steady state, as

$$F_1 = h(C^* - C_0), \quad F_2 = (D_{\text{eff}}/x)(C_0 - C_i),$$

$$F_3 = k_3 C_i. \tag{15}$$

Continuity of flux requires $F_1 = F_2 = F_3$, and we can solve for C_0 and for C_i, where C_0, C_i, and the other variables are the usual variables in the Deal–Grove derivation:

$$C_0 = \frac{xhC^*}{D_{\text{eff}} + xh} + \frac{D_{\text{eff}} C_i}{D_{\text{eff}} + xh}, \tag{16}$$

$$C_i = \frac{hD_{\text{eff}} C^*}{k_3 D_{\text{eff}} + hD_{\text{eff}} + xk_3 h} \approx \frac{D_{\text{eff}} C^*}{D_{\text{eff}} + k_3 x}, \tag{17}$$

since $h \gg k_3$. Substituting for C_i from Eq. (17) into Eq. (16) gives

$$C_0 = \frac{xhC^*}{D_{\text{eff}} + xh} + \frac{D_{\text{eff}}}{D_{\text{eff}} + xh} \frac{D_{\text{eff}} C^*}{D_{\text{eff}} + k_3 x}. \tag{18}$$

Values for k_3 and D_{eff} under typical thermal oxidation conditions have been previously determined. At 1000 °C and 760 Torr of oxygen, $k_3 = 3.6 \times 10^4$ μm/h and $D_{\text{eff}} = 2531$ μm^2/h[8]. Using these values, the following approximations may be made for oxidation under similar conditions:

$$C_0 \approx C^*, \quad C_i \approx \frac{D_{\text{eff}} C^*}{D_{\text{eff}} + k_3 x}. \tag{19}$$

Rewriting Eq. (13) using these forms for the concentrations of oxygen and reformulating the constants yield

$$Ndx/dt = k_3 C_i + \alpha C_0 e^{-x/a}, \tag{20}$$

and substituting for C_i and C_0 gives

$$Ndx/dt = \frac{k_3 D_{\text{eff}} C^*}{D + k_3 x} + \alpha C^* e^{-x/a}, \tag{21}$$

$$dx/dt = \frac{k_3 D_{\text{eff}} C^*}{N(D + k_3 x)} + \frac{\alpha C^* e^{-x/a}}{N}. \tag{22}$$

This expression for dx/dt may be compared with existing data on the oxidation rate of silicon for evaluation of the model. In the thin film limit (<20 nm), it should accurately predict the observed enhancement in oxide growth rates as compared with the rates at greater oxide thicknesses. Beyond oxide thicknesses of 20 nm, the relationship must give rates that are comparable with those predicted by the Deal–Grove model and with experimental observations. The parameter a in the exponential term should be a characteristic length which determines the onset of diffusional effects in the relationship.

IV. COMPARISON: MODEL VERSUS EXPERIMENT

The result of Eq. (22) was calculated and compared with experimental data in the literature. The most reliable data for thin oxidations, the region of interest, appears to be that of Massoud, Plummer, and Irene,[8] and comparison with these results was used as a benchmark in evaluating parameters in the model. It was found that the growth rates predicted by the model were strongly influenced by the ratio of k_3/α. With $k_3/\alpha < 1$, enhanced growth rates were calculated for the initial phase of oxidation with the model reducing to typical Deal–Grove values in the thick limit. Comparisons of growth rate versus thickness curves for our model, the model of Massoud, Plummer, and Irene, and Massoud's experimental data are presented in Fig. 4 for oxidation rates in the thin oxide regime. Our analytical model may be seen to predict enhanced growth rates in the early oxidation, with the interception point of various forms of the model determined by the value of a in the exponential term. Based on our derivation, a should represent some characteristic length within the analysis. Various tests for values of a were conducted in
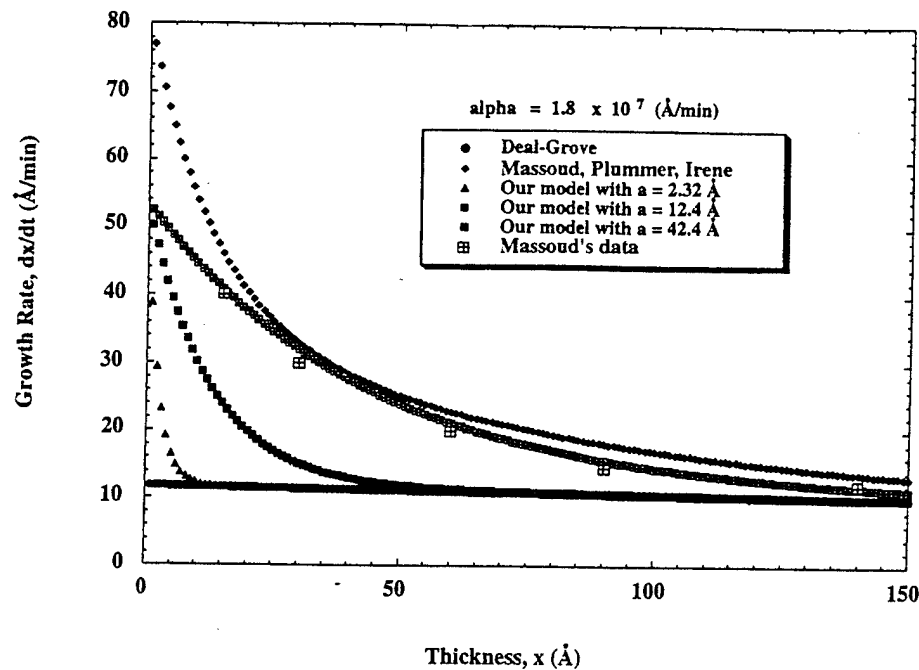
FIG. 4. Growth rate vs thickness relationships for silicon (100) dry oxidations at 1000 °C.

order to determine the best fit to the experimental data. It may be seen from Fig. 4 that an almost ideal fit of model to data is achieved with a value of $a = 4.24$ nm. We believe that this value denotes the oxide thickness at which diffusional limitations on the oxygen concentration at the interface begin to control the reaction.

V. AUGER MICROPROBE ANALYSES

The initial oxidation of a (2×1) silicon (100) surface at room temperature is studied with Auger electron spectroscopy. Cleaned surfaces are passivated in dilute HF and heated to 600 °C instead of the usual high-temperature sputter-clean/anneal cycles employed in many previous studies.[21-25] Water vapor in the oxygen gas at 1 atm pressure used for exposures is below 0.5 ppm. The effects of surface roughening and/or the introduction of water vapor on these surfaces is measured by comparison with sputter-cleaned surfaces and ambient air exposures. Hydrogen-desorbed silicon surfaces have less than 0.032 nm coverage after exposure to the ultradry oxygen at room temperature. The presence of water vapor allowed this to increase to 0.236 nm. It is possible to account for the presence of excess oxygen on the surface by using a calculation technique based upon decomposition of the silicon LVV Auger spectra.

Clean silicon (100) samples, p type of 5–10 Ω cm, were prepared by first using a 10:1 HF etch, with organic residuals removed in a 70:30 sulfuric acid:hydrogen peroxide solution followed by a 10 min dip in 1% HF in order to hydrogen-passivate the surface.[26,27] Samples are immediately loaded into an ultrahigh vacuum chamber of a PHI 600 Scanning Auger Multiprobe and heated to 600 °C for 30 min for hydrogen desorption and surface reconstruction.[21,28,29] Cooled

samples are exposed, in one case, to 1 atm ultrapure oxygen (99.9999%, <0.5 ppm water vapor) in the load lock of the Auger system at room temperature, then reinserted into the vacuum chamber through the load lock without any exposure to ambient air. In order to verify that surfaces are reactive, samples are, alternatively, exposed to moist ambient air. A part of the sample is sputter cleaned in order to compare the reactivity of smooth, hydrogen-desorbed surfaces to sputter-cleaned, roughened surfaces under the same conditions. In all cases, cleaned surfaces have no detectable oxygen or fluorine and negligible carbon before heating. After heating, there is some oxygen on the surface probably due to chamber or sample stage outgassing of water vapor, but it is below 2%. The vacuum remained at 10^{-10} Torr except for brief transients when power is applied maximally to the heating filament.

Figure 5 shows the measured silicon LVV Auger spectra for the HF-treated samples, while Fig. 6 depicts the sputter-cleaned samples under comparable conditions. The hydrogen-passivated line shape differs from the sputter-cleaned surface noticeably in that the former has a prominent peak (denoted a) at about 15.5 eV below the main peak at 89 eV. This is presumably the bulk plasmon loss peak which is also larger in the backscattered primary spectrum. This peak is reduced by heating and subsequent cooling or by heavy electron beam bombardment as would be expected for hydrogen desorption and a change in surface structure from (1×1) to (2×1). The Auger spectrum for the HF-treated sample exposed to ultradry oxygen for 60 min shows virtually no sign of oxidation, while the sputtered surface shows the predictable signs of early stage oxidation: a peak at 82 eV (denoted b) usually assigned to SiO_x and one at 76 eV (denoted c) for
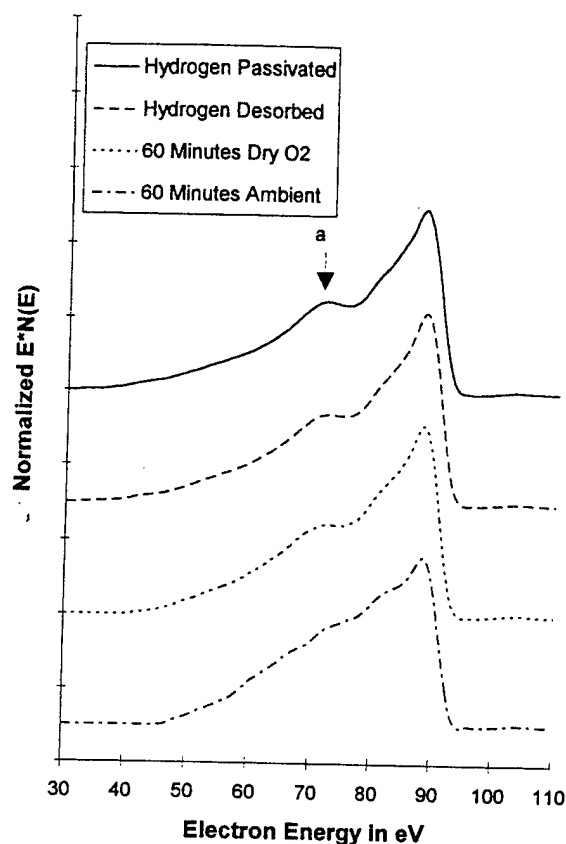
FIG. 5. Si *LVV* Auger spectra for HF-treated silicon (100) samples subjected to various oxidation process.

FIG. 6. Si *LVV* Auger spectra for sputter-cleaned silicon (100) surfaces subjected to varied oxidation processes.

SiO_2.[30,31] More activity is displayed in the spectrum of smooth hydrogen desorbed surfaces exposed to moist air.

Table I lists the results of the line shape analysis applied to quantify the level of oxide growth. It is found that, during various investigations of the desorbed (relatively inert) surface, oxygen can accumulate on the surface without affecting the silicon spectrum especially when exposure is interrupted for growth rate measurements. Therefore, exposures shown are not interrupted and oxide coverage is determined by the degree of oxidation shown in the silicon line shape itself.

Oxide coverage is determined by decomposition of the normalized background subtracted silicon spectrum into three parts: Si_{Si}, Si_{SiO_2}, and Si_{SiO}, where the sum of these three components is unity. The total count is then given by

$$total\ count = Si_{Si} + 3Si_{SiO_2} + 2Si_{SiO},$$

assuming the stoichiometry of SiO_x to be given by $x=1$. The fraction of silicon and silicon dioxide in the count is

$$F_{Si} = Si_{Si}/(total\ count),\quad F_{SiO_2} = 3Si_{SiO_2}/(total\ count).$$

Attenuation of the pure silicon component by oxide is then used to determine oxide thickness using the logarithmic decay due to inelastic scattering characterized by the escape depth parameter λ adjusted for detection angle.[32,33] Thus

$$thickness = \lambda_{ML}\ \ln(F_{Si}).$$

A value of $\lambda_{ML} = 3.13$ ML $= 0.726$ nm, where ML denotes monolayer, is used in the calculations based upon a quartz-
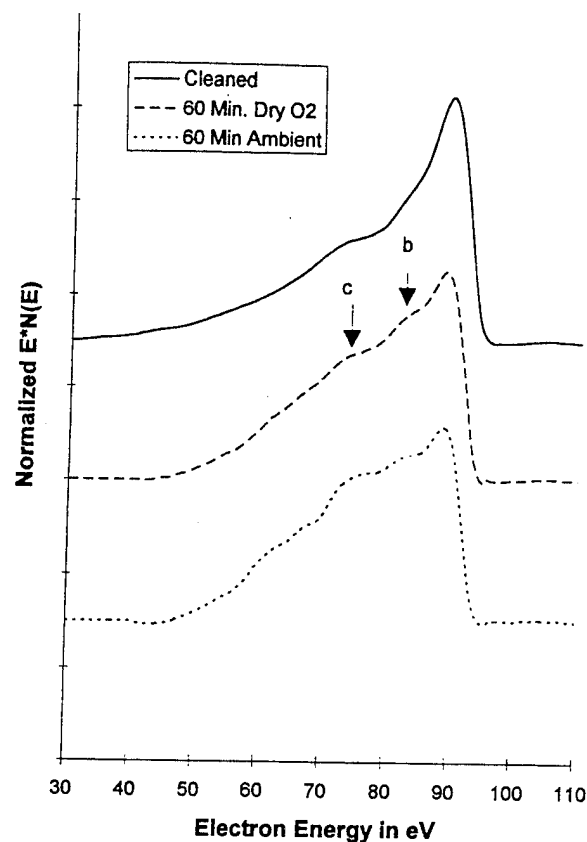
like density and a 30° detector angle. The predicted oxygen concentration is given by F_{SiO_2}. The reacted oxygen is simply the predicted oxygen concentration divided by the measured concentration. Predicted oxygen might be higher than actual because measured oxygen Auger electrons are less attenuated than the Si_{SiO} electrons. This technique is very sensitive in the 0–3 nm thickness regime, where small changes in coverage cause obvious attenuation of the underlying species.

The lack of growth for hydrogen-desorbed smooth Si(100) surfaces supports the assertions of previous work.[24,34,35] Although this technique is limited in absolute accuracy, there are a few clear interpretations: (a) clean, smooth Si(100) is unreactive in absolutely dry oxygen at room temperature even after heat treatments have supposedly removed the hydrogen passivation, and (b) growth is slight even with the help of water vapor. Finally, for the thinnest oxides (HF treated, dry O_2), there is little evidence for [SiO], as predicted by Eqs. (13) and (14). However, SiO formation is a fundamental process of our model, and these results suggest a strong tendency for island growth in the initial phases.

The use of a constant escape depth when the structure of the initial oxide is not completely understood is less than ideal. Work is continuing in this area, in the interpretation of Auger peaks, and in the improvement of background subtraction.

VI. CONCLUSIONS

We have analyzed the dry oxidation of silicon utilizing a model that invokes dissociative chemisorption in silicon at the interface between the silicon dioxide film and the substrate. A model is presented which supports the diffusing species in such processes to be molecular rather than atomic oxygen. The interfacial chemisorption framework as a model for silicon oxidation has been shown to account for current observations on the initial stages of silicon oxidation in ultradry oxygen, in that the model readily accommodates submonolayer reaction schemes in which the silicon dimer is retained on the surface. The model suggests a temperature dependence in the very early stages of the reaction that accounts for conflicting observations on the relative degree of silicon oxidation on exposure to oxygen at room temperature (RT) and at temperatures intermediate between RT and that required for sustained oxidation. The model predicts a self-limit of 0.5–0.6 nm in oxidations performed at temperatures sufficient to dissociate surface dimers and to permit oxygen penetration of the substrate beyond the first monolayer of suboxide. Bond energy considerations suggest that this temperature regime is probably in the 400–600 °C interval for dry oxidations. Detailed examination of the model has also suggested a mechanism which can account for the observation that oxide/silicon interfaces exhibit an inherent interfacial roughness of approximately one atomic diameter or about 0.3 nm.

Kinetic rate equations based on the model have been developed which successfully account for the observed power law dependence of rate on oxygen partial pressure. These relationships have been used in the derivation of an expression for the variation of oxide film growth rate with overlying oxide thickness. The relationship has been tested against experimental observations reported in the literature and found to give an excellent fit.

[1]S. Wolf and R. N. Tauber, in *Silicon Processing for the VLSI Era* (Lattice, Sunset Beach, CA, 1986), Vol. 1.

[2]S. M. Hu, J. Appl. Phys. **55**, 4095 (1984).

[3]B. E. Deal and A. S. Grove, J. Appl. Phys. **36**, 3770 (1965).

[4]L. N. Lie, R. R. Razouk, and B. E. Deal, J. Electrochem. Soc. **129**, 2828 (1982).

[5]Y. J. van der Meulen, J. Electrochem. Soc. **119**, 530 (1972).

[6]E. A. Irene, J. Electrochem. Soc. **125**, 1708 (1978).

[7]J. Blanc, Appl. Phys. Lett. **33**, 424 (1978).

[8]H. Z. Massoud, J. D. Plummer, and E. A. Irene, J. Electrochem. Soc. **132**, 2693 (1985).

[9]A. G. Revesz and R. J. Evans, J. Phys. Chem. Solids **30**, 551 (1969).

[10]E. A. Irene, J. Appl. Phys. **54**, 5416 (1983).

[11]J. M. Gibson and D. W. Dong, J. Electrochem. Soc. **127**, 2722 (1980).

[12]A. Fargeix, G. Ghibaudo, and G. Kamarinos, J. Appl. Phys. **54**, 2878 (1983).

[13]S. A. Ajuria, P. U. Kenkare, A. Nghiem, and T. C. Mele, J. Appl. Phys. **76**, 4618 (1994).

[14]B. J. Mrstik and P. J. McMarr, Phys. Rev. B **48**, 17972 (1993).

[15]A. Redondo, W. A. Goddard III, C. A. Swartz, and T. C. McGill, J. Vac. Sci. Technol. **19**, 498 (1981).

[16]A. Oshiyama and Y. Miyamoto, in *New Horizons in Low-Dimensional Electron Systems*, edited by H. Aoki *et al.* (Kluwer Academic, Amsterdam, 1992), pp. 337–350.

[17]X. M. Zheng and P. V. Smith, Surf. Sci. **232**, 6 (1990).

[18]H. Yaguchi, K. Fujita, S. Fukatsu, Y. Shiraki, R. Ito, T. Igarashi, and T. Hattori, Surf. Sci. **275**, 395 (1992).

[19]Y. Nakagawa, A. Ishitani, T. Takahagi, H. Kuroda, H. Tokumoto, M. Ono, and K. Kajimura, J. Vac. Sci. Technol. A **8**, 262 (1990).

[20]T. Ohmi, M. Morita, A. Teramoto, K. Makihara, and K. S. Tseng, Appl. Phys. Lett. **60**, 2126 (1992).

[21]H. Ibach and J. E. Rowe, Phys. Rev. B **9**, 1951 (1974).

[22]F. Lutz, L. Kubler, J. L. Bischoff, and D. Bolmont, Phys. Rev. B **40**, 11747 (1989).

[23]F. Lutz, J. L. Bischoff, L. Kubler, and D. Delmont, Phys. Rev. B **40**, 10356 (1989).

[24]V. D. Borman, E. P. Gusev, Yu. Yu. Lebedinski, and V. I. Troyan, Phys. Rev. B **49**, 5415 (1994).

[25]J. R. Engstrom, D. J. Bonser, M. M. Nelson, and T. Engel, Surf. Sci. **256**, 317 (1991).

[26]Y. Nakagawa, A. Ishitani, Takahagi, H. Kurodo, H. Tokumoto, M. Ono, and K. Kajimura, J. Vac. Sci. Technol A **8**, 262 (1990).

[27]M. Niwano, J.-I. Kageyama, K. Kurita, K. Kinashi, I. Takahashi, and N. Miyamoto, J. Appl. Phys. **4**, 2157 (1994).

[28]B. S. Meyerson, F. J. Himpsel, and K. J. Uram, Appl. Phys. Lett. **57**, 1034 (1990).

[29]H. Tomita, T. Kikuchi, and K. Furuya, Jpn. J. Appl. Phys. **30**, 897 (1991).

[30]T. Kunjunny and D. K. Ferry, Phys. Rev. B **24**, 4604 (1981).

[31]C. R. Helms, Y. E. Strausser, and W. E. Spicer, Appl. Phys. Lett. **33**, 767 (1978).

[32]M. P. Seah and W. A. Dench, Surf. Interface Anal. **1**, 1 (1979).

[33]D. Briggs and M. P. Seah, *Practical Surface Analysis* (Wiley, New York, 1990), Vol. 1.

[34]H. Yaguchi, K. Fujita, S. Fakatsu, Y. Shiraki, R. Ito, T. Igarashi, and T. Hattori, Surf. Sci. **275**, 395 (1992).

[35]M. Morita, T. Ohmi, E. Hasegawa, M. Kawakami, and M. Ohwada, J. Appl. Phys. **68**, 1272 (1990).

A UNIFIED CHEMICAL MODEL FOR THERMAL OXIDATION OF SILICON (100) IN A DRY OXYGEN AMBIENT

by

Prabhaharan Thanikasalam and David K. Ferry

Department of Electrical Engineering
Arizona State University
Tempe, AZ 85287

Introduction

Silicon dioxide is the most important component in the fabrication of a MOS device. In spite of the tremendous research in the last thirty years, it has not been possible to accurately model the initial oxidation phase of silicon [1-12]. Thermally grown silicon dioxide in VLSI processing application ranges in thickness from 6 nm -1000 nm. Some of the major functions of these films include a) masking against ion implantation and diffusion; b) passivation of silicon surface; c) isolation of individual devices; d) use as a gate oxide and capacitor dielectric in MOS devices; and e) use as a tunneling oxide in ROMs etc. [1]

The understanding of oxide formation mechanisms that was developed in the mid-1960's was inadequate for determining the process and material properties important to device performance as the device fabrication began to scale down into submicron regime. As device geometries have shrunk, a number of issues previously considered unimportant or unobserved, have been found to require a better resolution in order for deep sub-micron and nanoscale devices to be constructed [2].

Thermal oxidation of silicon has been usually modeled within the framework originally formulated by Deal and Grove [3]. They proposed that the oxidation rate was determined by a combination of two processes. The first involved the actual chemical reaction of oxygen with silicon at the oxide/substrate interface, while the second involved the diffusion of oxygen through the already formed oxide film. The combination of these processes resulted in the formulation of the classical "linear-parabolic" rate law of Deal-Grove. However, the simple linear-parabolic rate law could not accurately explain the growth curves obtained for oxides of thicknesses below 20 nm. Massoud and Plummer [4] proposed a modification to the original Deal-Grove model by the addition of two separate exponential terms, although they did not associate any relevant physical processes that would be necessary to justify the addition of these two exponential terms. However, they claimed that this fit the data better in the thin oxide regime.

Background

It has been shown by experiments that oxidation proceeds by the inward movement of a species of oxidant rather than by the outward movement of silicon [2,3,4,5]. Based on this concept, Deal and Grove proposed their model:

$$\frac{dx}{dt} = \frac{F}{N_1} = \frac{1}{N_1}\frac{kC^*}{\left[1+\dfrac{k}{h}+\dfrac{kx}{D_{eff}}\right]} \quad ,$$

where x is the oxide thickness, F the total flux of oxidant molecules through the oxide, k the first order rate constant for oxidation, C^* the equilibrium concentration of oxidant in the gas phase, N_1 the number of oxidant molecules incorporated into a unit volume of the oxide layer, h the gas phase mass transfer coefficient of oxygen and D_{eff} the effective diffusion coefficient of oxygen in silicon dioxide. Solution of the differential equation will result in

$$x^2 + Ax = B(t + \tau) \quad ,$$

where

$$A = 2D_{eff}\left(\frac{1}{k}+\frac{1}{h}\right) \quad ,$$

$$B = 2D_{eff}\frac{C^*}{N_1} \quad ,$$

and

$$\tau = \frac{\left(x_0^2 + Ax_0\right)}{B} \quad .$$

The Deal-Grove model requires the parameter τ (a shift in the time co-ordinate) in order to account for the presence of an initial oxide layer on the silicon surface.

The above approach was found to provide an excellent fit to experimental data for oxidation processes utilizing H2O/O2 mixtures and for dry oxidation processes at thicknesses in excess of ca. 40 nm [6]. Deal-Grove estimates fail, however, in predicting oxidation rates in dry oxygen within the thin regime(< 20 nm) and in accounting for the observed pressure dependence of both thin and thick oxidations. Within the thin regime, observed oxidation rates in dry processes are significantly greater than the Deal-Grove predictions and the reaction exhibits a power law dependence on the oxygen pressure with rate propotional to P^m, where m=0.6-0.8 [7,8,9].

Space-charge effects have been postulated to give enhanced rates in thin oxidations [9]. This phenomenon requires the oxidant to exist and diffuse in *ionic* form and is operative only over the oxide thicknesses comparable to the extent of the space-charge region. Based on detailed analysis of experimental parameters, oxidation results, and the effects of the electric field at the interface on oxygen transport, it generally has been concluded that space charge effects are not the mechanism giving rise to enhanced oxidation rates within the thin oxidation regime [4].

Other attempts at explaining elevated rates in the thin oxidations proposed structural differences between thin and thick oxides, which lead to enhanced oxygen diffusion in the thin region [10]. The presence of microchannels with diameters of about 5 nm has been proposed to increase oxygen transport in thin oxides [11]. Some experimental evidence for such microchannels exists [12]. Stress at the oxide/substrate interface is also thought to affect the rate of oxidation [11]. Stress has been invoked to vary either the transport of oxygen through the oxide film or the intrinsic reaction rate at the interface [13-15]. Finally, variations in the oxidizing species have been suggested as the potential sources of variation in oxidation rates in the thin and thick regimes [16]. Massoud's modification of the total growth rate expression is given by

$$\frac{dx}{dt} = \frac{B}{(2x+A)} + C_1 e^{-\frac{t}{L_1}} + C_2 e^{-\frac{t}{L_2}} \quad .$$

The pre-factor C_1 and exponent L_1 describes the initial phase of the growth and C_2 and L_2 describe the intermediate phase of the growth. Jorgensen [17] suggested that the diffusing species in dry

oxygen is a singly ionized oxygen molecule, O_2^-. However, the experiments of Jorgensen are subject to objections of Modlin's [18] experiment, which gave results opposite to that of Jorgensen. Furthermore, Modlin's results indicated that, for diffusion-limited thermal oxidation, the diffusing species is neutral molecular oxygen.

Most of the above discussions retain the Deal-Grove description. They also retain the view that only the standard diffusion mechanism occurs (e.g. oxygen molecules diffuse interstitially), and then they introduce some modest change, like a stress gradient, and hence have a weak dependence of the diffusion and interface reaction parameters in the Deal-Grove expressions. The Mott-Cabrera [19] model relies upon two key ideas. First, there is an electron tunneling process which tranfers charge across the oxide until the potential generated prevents further transfers. Secondly, the electric field set up in this process affects the injection of mobile ions into the oxide; there are no effects on the process of transport across the oxide in this model.

It is widely believed (and there is overwhelming evidence [22, and references therein]) that most transport is by neutral species, but there are special cases (e.g. in the presence of an electron beam) when charged species may be important. Some workers strongly favor charged species. Some, even though supporting the neutral oxygen molecule transport in the thick regime, believe that the moving oxygen species in the thin oxide growth regime is charged [21].

Proposed Silicon Oxidation Mechanism

The conceptual model proposed here involves dissociative chemisorption of molecular oxygen for the process of oxidation of silicon [32]. To illustrate the model, we start from an ideal clean (2×1) reconstructed Si (100) surface [See Figs. 2 & 3]. Sphere sizes reflect the relative covalent radii of silicon and oxygen atoms. The system always tries to minimize the surface energy by minimizing the number of dangling bonds on the surface or by satisfying as many dangling bonds as possible. It is assumed that the surface has minimized energy by reconstruction. After a (2×1) reconstruction, the number of dangling bonds is reduced to one half the unreconstructed

surface value, which still gives rise to a number of different adsorption sites on the reconstructed surface, which is apparent in the diagram. The nature of these adsorption sites and the energetic studies for oxygen adsorption have been reported by Oshiyama [22] and Zheng [5].

The molecular orbital energy diagram for ground state oxygen is shown in Fig. 4(a). The electronic configuration of the oxygen atom is explained in this section. The Highest Occupied Molecular Orbital (HOMO) consists of two half-filled orthogonal and degenerate orbitals of anti-bonding symmetry. The 95% probability surfaces for the orbitals are shown in Fig. 4(b). The mixing of the anti-bonding HOMO with the highest energy orbitals of the silicon surface initiates the very first interactions between the incoming oxygen molecule and the silicon substrate. The HOMO on the (2×1) reconstructed Si (100) surface is postulated to be a half-filled "dangling bond" on each silicon atom in the dimer units, presumably of sp^3 hybridized symmetry. Relative geometries of the surface "dangling bond", projecting from the dimer into the gap between dimer rows, and the orthogonal, half-filled π^* anti-bonding HOMO on the oxygen molecule suggest that maximum orbital mixing will occur when adsorption takes place in the initial oxygen chemisorption step illustrated in Fig. 2. The interacting anti-bonding HOMO on the oxygen molecule are π^* in nature. Therefore, the electronic density incorporation from the surface silicon acts to weaken the oxygen-oxygen bond, leading to a dissociation process of the oxygen molecule and a consecutive insertion of the oxygen atoms into neighboring silicon-silicon backbonds. At this stage it is not clear whether these three processes of chemisorption, bond dissociation and bond insertion occur as a concerted process or in discrete steps.

Depending on the growth temperature, three different initial reactions can be proposed under anhydrous conditions, based on experimental evidence and surface analytical studies done in our laboratories and other facilities [23]. Figure 2 shows a possible reaction scheme for the dissociation of molecular oxygen on the silicon surface at room temperature. Even though thermodynamically favored Si-O bond formation may occur at lower temperatures, there appears to be insufficient energy for dissociation of the silicon dimer bonds and the reaction may be kinetically limited to the formation of oxygen bridges between dimers with oxygen coverage of ca.

0.5 monolayer [23]. However, as the temperature of the process is increased, the availability of the energy required for the dissociation of the surface dimer bonds would increase thus enhancing the probability of the surface to evolve from a (2×1) to (1×1) oxygen surface superlattice through extensive oxygen coverage. This produces a single monolayer coverage on the top surface. However, there is not much experimental evidence supporting such structure formation [24,25].

Oxidation studies performed at atmospheric pressure conditions and at suitably higher temperatures, in order to sustain oxidation, apparently appear to self-limit at thicknesses of less than 1 nm, whereas typical reported self-limiting thicknesses range from 0.6 - 1.5 nm [4]. Fig. 3 illustrates a possible explanation to this. Layer-by-layer growth of the very initial regime could be explained by a mechanism which involves interactive action of molecular oxygen chemisorption at the oxide-silicon interface followed by molecular dissociation of oxygen and insertion into silicon backbonds. The extent of this mechanism for the initial layer-by-layer growth is crucially dependent on the chemisorption of molecular oxygen through orbital mixing with the unoxidized silicon at the silicon/oxide interface. From Fig. 3, it is quite clear that the unavailability of the unoxidized silicon to the oxygen adsorbed on the surface limits the initial layer-by-layer growth without an additional driving force to drive the oxygen through the thin oxide film. This kind of masking or barrier effect is produced by just two fully oxidized layers of silicon and the reaction will therefore cease at this juncture for any temperature below those at which appreciable oxygen diffusion to the interface can occur. The approximate thickness of this self-limiting layer of oxide film is 0.5 - 0.6 nm, which is consistent with experimentally reported values for self-limiting oxide films [4, and references therein].

This model adequately describes the structural phenomena reported for dry oxidation of clean silicon (100) surfaces. However, it should be noted that this model does not address oxidation processes in the presence of moisture and it does not address native oxide growth or native oxide thicknesses that result from exposure to a moist ambient prior to dry oxidation. It is expected that the presence of moisture would significantly change the basic physical and chemical

mechanisms at the very fundamental level and hence cannot be compared with dry ambient experiments.

This model also addresses the origin and evolution of the inherent interfacial roughness which is always present at the silicon/oxide interface and which has been characterized to extend to ca. one atomic diameter. Fig. 5 shows a (011) projection of the silicon (100) surface during the process of oxidation. At the top end of the figure, a mixed configuration of the initial oxide layer is shown where both the (2×1) and (1×1) bridged oxygen structures are shown on the surface. A transient temperature ramping during the oxidation process may produce such a mixed configuration on the surface. Such a surface configuration is expected to be the cause of the origin of the inherent interfacial roughness. Regions of the surface that have retained the (2×1) reconstructed structure would effectively cause a temporary blocking of oxygen penetration and insertion phenomena in the underlying silicon backbonds. Penetration and insertion of the oxygen molecule and oxidation may proceed in the areas with the more open (1×1) reconstructed arrangement. Oxidation of the top layer back bonds gives a combination of the two types of surface structures as shown in the second and third schematics of Fig. 5. As the oxidation proceeds, and the dimer is eventually oxidized, the discrepancy between the oxidation depth under initially (1×1) and (2×1) surfaces may be expected to propagate with the interface, giving rise to an inherent interfacial roughness of 0.3 nm, approximately the diameter of an Si-O grouping. Oxidation to the final condition exhibiting such interfacial roughening is depicted in the lower three schematics of Fig. 5.

Silicon Oxidation Kinetics

The conceptual model, described in the previous section, is utilized to formulate the rate equations that adequately describe the oxidation mechanism of silicon (100) surface in a dry oxygen ambient. According to the model, the fundamental chemical reaction occurs via classical Langmuir kinetics. A major difference in this model when compared to existing models is that there

is a production of an intermediate species prior to the final reaction product SiO$_2$. At this juncture, this intermediate product is termed "SiO". Even though several experimental results have confirmed that the average stoichiometry of the structure that exists in the vicinity of the silicon/oxide interface corresponds to SiO, it is acknowledged that a completely rigorous treatment would require a more extensive experimental and theoretical analysis of the intermediate product and subsequently more detailed series of equations to describe the chemisorption process at the interface. Adopting the Langmuir kinetic formalism, the chemical reaction sequence for the oxidation of silicon may be written as:

$$Si + \frac{1}{2}O_2 \underset{k_1^{'}}{\overset{k_1}{\Longleftrightarrow}} SiO \quad , \tag{1}$$

$$SiO + \frac{1}{2}O_2 \overset{k_2}{\longrightarrow} SiO_2 \quad , \tag{2}$$

It should be noted that the first reaction is reversible (i.e. in equilibrium), but the second is not. The existing equilibrium between the molecular oxygen chemisorbed at the silicon/oxide interface and the substrate silicon is represented by the first reaction sequence, whereas the second reaction sequence describes the non-equilibrium reaction between the intermediate species and the additional chemisorbed oxygen to yield the final reaction product, silicon dioxide, SiO$_2$. $k_1, k_1^{'}$ and k_2 are the usual reaction rate constants.

Using the second reaction, the rate of formation of the final product, SiO$_2$, may be expressed as,

$$\frac{d[SiO_2]}{dt} = k_2[SiO][O_2]^{\frac{1}{2}} \quad , \tag{3}$$

where [O$_2$] is the concentration of chemisorbed molecular oxygen available at the silicon/oxide interface and [SiO] is the concentration of the partially oxidized intermediate species at the silicon/oxide interface. The concentration of oxygen at the interface can be determined by

employing the traditional diffusional approaches similar to those applied in Deal-Grove model and most other current reaction models. The availability of oxygen in the equilibrium reaction will vary within the reaction time frame due to diffusion limitations. Therefore any steady-state assumptions would not be valid in this reaction and the usual equilibrium relationships will not be considered. However, absolute expressions for reaction rates remain valid and the concentration of the intermediate species may be calculated by

$$\frac{d[SiO]}{dt} = k_1[Si][O_2]^{\frac{1}{2}} - k_1'[SiO] - k_2[SiO][O_2]^{\frac{1}{2}} \qquad . \qquad (4)$$

After the formation of a thick layer of oxidized silicon, the concentration of the intermediate product SiO remains almost a constant. In other words, the rate of change of [SiO] is zero. i.e. d[SiO]/dt = 0 in the long-time limit. In the long-time limit, (4) gives

$$[SiO] = \frac{k_1[Si][O_2]^{\frac{1}{2}}}{k_1' + k_2[O_2]^{\frac{1}{2}}} \qquad . \qquad (5)$$

Substituting this into (3), the rate of formation of silicon dioxide film can be expressed as

$$\frac{d[SiO_2]}{dt} = \frac{k_1 k_2[Si][O_2]}{k_1' + k_2[O_2]^{\frac{1}{2}}} \qquad . \qquad (6)$$

Let N be a conversion factor for film thickness to concentration of SiO_2, analogous to that employed in the Deal-Grove model. The concentration of silicon [Si] is a constant at the oxide/silicon interface. The growth rate can be rewritten as,

$$N\frac{dx}{dt} = \frac{k_1 k_2[Si][O_2]}{k_1' + k_2[O_2]^{\frac{1}{2}}} \qquad . \qquad (7)$$

It is clear that the oxide film growth rate is solely controlled by the availability of oxygen at the oxide/silicon interface. Therefore, the only crucial variable quantity as the growth time progresses is the concentration of oxygen $[O_2]$. The growth rate expression (7) was obtained from the fundamental reaction rate equation, (6), and it's applicable to the oxidation of silicon in a dry oxygen ambient. However, this would characterize the growth only in the thick oxide regime. Therefore, it is precisely this function that has to be modified for the thin transient regime (for thin oxides).

In order to get to the transient regime, (4) is multiplied on both sides by the integrating factor $e^{\gamma t}$ and the following expression is obtained:

$$\frac{d[SiO]}{dt}e^{\gamma t} = k_1[Si][O_2]^{\frac{1}{2}}e^{\gamma t} - \left(k_1' + k_2[O_2]^{\frac{1}{2}}\right)[SiO]e^{\gamma t} \quad . \tag{8}$$

If we define $\gamma \equiv k_1' + k_2[O_2]^{\frac{1}{2}}$, then we can rewrite (8) as

$$\frac{d[SiO]}{dt}e^{\gamma t} = k_1[Si][O_2]^{\frac{1}{2}}e^{\gamma t} - \gamma[SiO]e^{\gamma t} \quad . \tag{9}$$

Integrating both sides of (9) gives

$$[SiO] = \frac{k_1[Si][O_2]^{\frac{1}{2}}}{k_1' + k_2[O_2]^{\frac{1}{2}}}\left(1 - e^{-\gamma t}\right) \quad . \tag{10}$$

To obtain the oxide film growth rate in the transient thin oxide regime, we substitute this expression for $[SiO]$ into (3) to get

$$\frac{d[SiO_2]}{dt} = \frac{k_1 k_2[Si][O_2]}{k_1' + k_2[O_2]^{\frac{1}{2}}}\left(1 - e^{-\gamma t}\right) \quad . \tag{11}$$

According to (11), there is no production of SiO$_2$ for thin oxides. i.e. for small times. This is a very important factor which might suggest nucleation and island formation in the very initial stages. It is convenient to move to the spatial domain from the temporal domain at this point. Therefore, set $\gamma t = x/a$, where "a" is a characteristic length and "x" is the oxide film thickness and "t" is the time oxidation. This allows the growth rate to be rewritten as

$$N\frac{dx}{dt} = \frac{k_1 k_2 [Si][O_2]}{k_1 + k_2 [O_2]^{1/2}}\left(1 + e^{-x/a}\right) \qquad . \tag{12}$$

Equation (12) relates the growth rate of the oxide film to the concentration of molecular oxygen at the oxide/silicon interface. The concentration of molecular oxygen at the interface cannot be experimentally determined. Therefore it is necessary to derive a relationship between the interfacial molecular oxygen concentration and the surface concentration of oxygen molecules (i.e. at the gas/oxide boundary). This quantity can be calculated if the oxygen partial pressure in the gas phase is known. Employing Fick's law, along with Henry's law near the gas/oxide boundary, and invoking a flux picture in a derivation analogous to that of Deal-Grove, it is possible to obtain a relationship that relates concentrations of oxygen molecules near the gas/oxide and the oxide/substrate interfaces [Figure 1]. The steady-state fluxes are given by

$$
\begin{aligned}
F_1 &= h(C^* - C_o), \\
F_2 &= \frac{D_{eff}}{x}(C_o - C_i), \\
F_3 &= k_3 C_i.
\end{aligned}
\tag{13}
$$

The steady-state condition dictates continuity of fluxes. Therefore $F_1=F_2=F_3$, and solving for C_o and C_i in terms of the other variables D_{eff}, h, C^* and x, we obtain

$$C_o = \frac{hD_{eff}C^*}{D_{eff} + xh} + \frac{D_{eff}}{D_{eff} + xh}C_i, \tag{14}$$

$$C_i = \frac{hD_{eff}C^*}{k_3D_{eff} + hD_{eff} + xk_3h} \cong \frac{D_{eff}C^*}{D_{eff} + k_3x}, \tag{15}$$

since $h \gg k_3$. Substituting for C_i from (15) into (14) gives

$$C_o = \frac{hD_{eff}C^*}{D_{eff} + xh} + \frac{D_{eff}}{D_{eff} + xh}\left(\frac{D_{eff}C^*}{D_{eff} + k_3x}\right). \tag{16}$$

Values of k_3 and D_{eff} under typical thermal oxidation conditions have been previously determined. At 1000°C and 760 Torr. of oxygen, $k_3 = 3.6 \times 10^4$ μm/hr and $D_{eff} = 2531$ μm^2/hr [26]. Using these values, the following approximations may be made for oxidation under similar conditions:

$$C_o \cong C^* \tag{17}$$

and

$$C_i \cong \frac{D_{eff}C^*}{D_{eff} + k_3x}. \tag{18}$$

Rewriting (12) using these forms for the concentrations of oxygen and reformulating the constants, yields

$$N\frac{dx}{dt} = k_3C_i + \alpha C_o e^{-x/a}, \tag{19}$$

and substituting for C_i and C_o gives

$$N\frac{dx}{dt} = k_3 \frac{D_{eff}C^*}{D_{eff} + k_3x} + \alpha C^* e^{-x/a}, \tag{20}$$

where α is the initial 'ballistic' reaction rate coefficient which is explained in detail later. Thus,

$$\frac{dx}{dt} = \frac{k_3D_{eff}C^*}{N[D_{eff} + k_3x]} + \frac{\alpha C^*}{N}e^{-x/a}. \tag{21}$$

This expression for dx/dt may be compared with existing experimental data on the oxidation rate of silicon for the evaluation of this model. In the thin limit (< 20 nm), this accurately predicts the observed enhancement in oxide growth rates as compared with the rates at larger oxide thickness. Beyond oxide thickensses of 20 nm, the relationship gives rates that are comparable with those predicted by the Deal-Grove model and with experimental observations. The parameter "a" in the exponential term is a characteristic length which is associated with the onset of diffusional effects of oxygen through the oxide as the film grows.

In the thin regime (for smaller values of "x"), the Deal-Grove model reduces down to

$$x \cong \frac{k_3 C^*}{N} t. \tag{22}$$

On the other hand, (21) reduces to

$$\frac{k_3 D_{eff} C^*}{N} dt \cong \frac{k_3 D_{eff}}{\alpha} e^{-x/a} dx, \tag{23}$$

$$x \cong \frac{\alpha C^*}{N} t. \tag{24}$$

Equation (24) gives an enhanced growth rate if $\frac{k_3}{\alpha} < 1$. In other words, the initial oxidation gives enhanced growth rate (for $x \to 0$) if

$$\alpha C_o > k_3 C_i$$

or

$$\frac{\alpha}{k_3} > 1 \ . \tag{25}$$

Results and Discussion

We use data taken from Massoud *et al.* [4], Irene and Van de Meulen [8], Kamigaki and Itoh [29] and Chao *et al.* [30]. As described above, this data is initially fit in the thick oxide regime with Deal-Grove. Fig. 6 shows the linear reaction coefficient k_3 and the diffusion coefficient D for all the data on a single plot. The activation energies associated with these two parameters are 1.54 eV and 2.58 eV, respectively. The Si-Si bond energy is 1.83 eV [21] and the closeness of the activation energy for k_3 to this value suggests that the controlling mechanism is the breaking of he Si-Si bonds. A higher activation energy in the diffusion-controlled regime explains the decreased growth rate in that regime where the growth rate is controlled by the diffusivity of the oxidizing species. As the oxidation temperature is raised, the oxidant molecules are driven deeper into the already formed oxide before they react at the interface with the substrate silicon. Raising of the temperature would produce substantial thermal vibration to open up the oxide network structure, thus enabling the oxidant to traverse further through the already formed oxide. It has been reported previously [3] that the activation energy associated with the bulk diffusion of oxygen in fused silica is 1.17 eV. This is far smaller than the value found here.

Massoud's [4] samples contained ca. 1.0 nm of oxide prior to thermal oxidation whereas Irene [8] reported the presence of 0.3 - 0.6 nm of oxide before the wafers were introduced into the oxidation chamber. Kamigaki's [29] samples reportedly had a 0.6 nm thick oxide layer while Chao's [30] ellipsometry detected a 1.6 - 2.0 nm oxide film prior to oxidation. Massoud, Irene and Chao used an *in-situ* ellipsometer whereas Kamigaki's samples were taken out of the oxidation furnace in order to perform the oxide thickness measurements by ellipsometer. The reported initial oxide thicknesses agreed quite well with the initial oxide thicknesses found in fitting the Deal-Grove model (as explained in the section above) to determine the k_3 and D values. All the samples were lightly doped. All the experiments were performed at atmospheric pressure. Following the first evaluation, (21) is now used to fit each set of data, but with k_3 and D evaluated from the straight lines in Fig. 6. The only variable is the parameter a.

Figure 7 shows the oxide thickness as a function of oxidation time for Massoud's experimental data and the fits to the model (21). The best value of the characteristic length a is found at each oxidation temperature. The value of the characteristic length a varies between 0.28 nm and 1.0 nm over the temperatures 800 - 1000 C. In Fig. 8, we show the comparison between model (21) and Irene's experimental data. The fit is excellent in all temperature ranges. The range of the characteristic length a is 0.3 - 1.0 nm. Kamigaki's experiment was done at several different pressure ranges. However, the comparison of Fig. 9 shows only the growth curves for the experiments performed under atmospheric pressure conditions. A minimal number of data points in Kamigaki's and Chao's experiments prevented us from making a definitive distinction between reaction controlled and diffusion controlled regimes for computation of the activation energies. The value of a again ranged from 0.15 to 0.30 nm for temperature ranges of 950 - 1100 C in Kamigaki's experiments, while we find 0.33 - 1.25 nm in Chao's experiments (Fig. 10). The fits are excellent.

In Fig. 11, we plot the values of the characteristic length a as a function of the growth temperature. The average value of a is ca. 0.35 nm, which is approximately the height of a Si-O molecule at the Si/SiO$_2$ interface for the temperatures up to 950 C. However, Kamigaki's data is different. In the latter data, nitrogen was used to dilute the oxygen gas. There is a good possibility that the nitrogen molecules would have initially blocked some of the surface sites, preventing a rapid initial growth regime.

The reaction controlled regime is a strong function of the initial surface structure prior to thermal oxidation. The cleaning processes, surface passivation techniques and the initial native oxide layer thickness are some of the major factors that would affect the value of the characteristic length at the very fundamental level. Significantly different cleaning and surface passivation techniques were used by the various researchers whose experimental data were compared in this report. They also had considerably different thicknesses of native oxide layer present on their samples prior to thermal oxidation process. Nevertheless, the values of a agree quite well.

The rise in a above 950 C presumably signals a more reactive interface process or a structural change or both. We postulate three different reactions that may occur at the very beginning of the oxidation process. The parameter a may very well characterise the thickness of the initial self-limiting oxide layer. At very low oxidation temperatures, the value of a is ca. 0.33 nm which agrees well with the idea that the top layer silicon dangling bonds are the only ones that are oxidized. As the temperature of oxidation is raised the oxygen molecules may have enough kinetic energy to penetrate through the silicon lattice to oxidize the back bonds and account for the rise in a. This may be true until the temperature is ca. 950 C. It has been reported by Irene, Tierney and Angiello [33] that there is a change in the density of the oxide film at ca. 950 C. They have observed a 3% change in density going from 600 C to 1150 C. Based on the values reported by Irene *et al.* [33], 2.28 - 2.20 gm/cm^3, comparisons with available data [23] indicate the formation of either Tridymite or Cristobalite forms of silica. We believe [34] that this change in the value of a may be related to the penetration depth of the initial layer of oxide formation which in turn depends on the structural porosity of the silicon lattice for oxygen. Higher oxidation temperatures may also enhance the porosity of the silicon lattice by opening up the top few layers thereby letting the initial oxidation to continue until ca. 1.2 nm of oxide is formed. However, at this stage it is not possible to conclude what form of silica is being formed at these temperatures and whether there is actually a structural change occuring at 950 C without further experimental evidence.

The fitting parameter α is kept constant so that $\alpha/k_3 > 1$. This parameter describes the rate at which the oxygen molecules penetrate the Si 2×1 reconstructed surface to oxidize the upper most layers of the substrate. A typical value [34] of α in the temperature range of 800 - 1000 C is ca. ~ 10^6 μ/min., whereas the linear reaction rate coefficient, k_3, is ca. ~ 10^3 μ/min. We expect α to have a stronger dependence on gas phase oxygen pressure rather than the oxidation temperature. Apparently, in this very initial *ballsitic* growth regime, for the lack of a better word, the diffusivity of oxygen species does not play a major role.

Conclusion

A model has been proposed for dry oxidation of silicon which invokes dissociative chemisorption in silicon at the interface between the silicon dioxide film and the substrate. This model supports the diffusing species in such processes to be molecular rather than atomic oxygen. This model also predicts a self-limiting thickness of 0.5 - 0.6 nm in oxidations performed at temperatures sufficient to dissociate surface dimers and permit oxygen penetration of the substrate beyond the first monolayer of sub-oxide. A preliminary examination of the model has also suggested a mechanism which can account for the observation that oxide/silicon interfaces exhibit an inherent interfacial roughness of approximately one atomic diameter or about 0.3 nm Kinetic rate equations have been developed and tested against experimental observation reported in literature and found to give an excellent fit. .

We have compared relevant experimental data for dry oxidation of silicon with a model that invokes dissociative chemisorption in silicon at the interface between the silicon dioxide film and the substrate. This relationship has been tested against significant experimental observations reported in the literature and found to give an excellent fit. It should be noted that insufficient data points in Kamigaki's and Chao's experiments prevented us from obtaining definitive activation energies. Different cleaning procedures and surface passivation techniques prior to oxidation mayalso be the cause of the significant deviation of the model from experimental observation in some of the cases. Dilution of the oxygen environment by nitrogen is postulated to play a crucial role at the very initial stages of oxidation by blocking some of the surface sites preventing a rapid initial growth. The value of a is believed to be the thickness of the very initial layer oxide that is formed when the 2×1 reconstructed silicon surface is exposed to an oxygen ambient. This might be a phenomena to be analyzed in detail experimentally in order to study island growth modes initiated at the very initial stages by blocked surface sites. It is our interpretation that the initial ballistic growth regime may be strongly dependent on oxygen pressure rather than on the temperature.

Further experimental evidence is required before conslusive remarks could be made on the structural evolution of the initial oxidation phase and the pressure dependence of the ballistic growth regime.

References

1. S. Wolf and R. N. Tauber, in *Silicon Processing for the VLSI Era*- **Vol. 1**, Lattice Press, Sunset Beach, CA (1986).

2. S. M. Hu, *J. Appl. Phy.*, **55**, 4095 (1984).

3. B. E. Deal and A. S. Grove, *J. Appl. Phy.*, **36**, 3770 (1965).

4. H. Z. Massoud, J.D. Plummer and E.A. Irene, *J. Elec. Chem. Soc.*, **132**, 2693 (1985).

5. X. M. Zheng and P.V. Smith, *Surf. Sci.*, **232**, 6 (1990).

6. L. N. Lie, R.R. Razouk and B.E.Deal, *J. Elec. Chem. Soc.*, **129**, 2828 (1982).

7. Y. J. van der Meulen, *J. Elec. Chem. Soc.*, **119**, 530, (1972).

8. E. A. Irene, *J. Elec. Chem. Soc.*, **125**, 1708, (1978).

9. J. Blanc, *Appl. Phy. Lett.*, **33**, 424 (1978).

10. A. G. Revesz and R.J. Evans, *J. Phy. Chem. Sol.*, **30**, 551 (1969).

11. E. A. Irene, *J. Appl. Phy.*, **54**, 5416 (1983).

12. J. M. Gibson and D.W. Dong, *J. Elec. Chem. Soc.*, **127**, 2722 (1980).

13. A. Fargeix, G. Ghibaudo and G. Kamarinos, *J. Appl. Phy.*, **54**, 2878 (1983).

14. S. A. Ajuria, P.U. Kenkare, A. Nghiem and T.C. McGill, *J. Appl. Phy.*, **76**, 4618 (1994).

15. B. J. Brstik and P.J McMarr, *Phy. Rev. B.*, **48**, 17972 (1993).

16. A. Redondo, W.A. Goddard III, C.A. Swartz and T.C. McGill, *J. Vac., Sci. Tech.*, **19**, 498 (1981).

17. P. J. Joegensen, *J. Chem. Phy.*, **37**, 874 (1962).

18. D. N. Modlin *Ph.D. Thesis*, Stanford University, Stanford, CA (1983).

19. N. Cabrera and N. F. Mott, *Rep. Prog. Phy.*, **12**, 163 (1949).

20. M. L. Yu et al., *Surf. Sci. Rep.*, **19**, 7 (1994).

21. C. J. Sofield and A.M. Stoneham, *Semi. Sci. Tech.*, **10**, 215 (1995).

22. A. Oshiyama and Y. Miyamoto in *New Horizons in Low-dimensional Electron Systems*, Kluwer Academic Publishers, Amsterdam (1992).

23. H. Yaguchi, K. Fujita, S. Fukatsu et al., *Surf. Sci.*, **275**, 395 (1992).

24. Y. Nakagawa, A. Ishitani et al., *J. Vac. Sci. Tech.*, **A8**, 262 (1990).

25. T. Ohmi, M. Morita, A. Teramoto et al., *Appl. Phy. Lett.*, **60**, 2126 (1992).

26. F. P. Fehlner , in *Low-Temperature Oxidation*, Wiley-Interscience Publication (1986).

27. E. A. Irene and Y.J. van der Meulen, *J. Elec. Chem. Soc.*, **123**, 1380 (1976).

28. M. A. Hopper, R.A. Clarke and L. Young, *J. Elec. Chem. Soc.*, **122**, 1216 (1975).

29. Y. Kamigaki and Y. Itoh, *J. Appl. Phy.*, **48**, 2891 (1977).

30. T. S. Chao, C.L. Lee and T.F. Lei, *J. Elec. Chem. Soc.*, **138**, 1756 (1991).

31. K. K. Ng, W. J. Polito, J.R. Ligenza, *Appl. Phy. Lett.*, **44**, 626 (1984).

32. T. K. Whidden, P. Thanikasalam, M.J. Rack and D. K. Ferry, *J. Vac. Sci. Tech.* **B 13**, 1618 (1995).

33. E. A. Irene, E. Tierney, J. Angiello. *J. Elec. Chem. Soc.*, **129**, 2594 (1982).

34. P. Thanikasalam, T. K. Whidden and D. K. Ferry, *J. Vac. Sci. Tech.* (to be published).
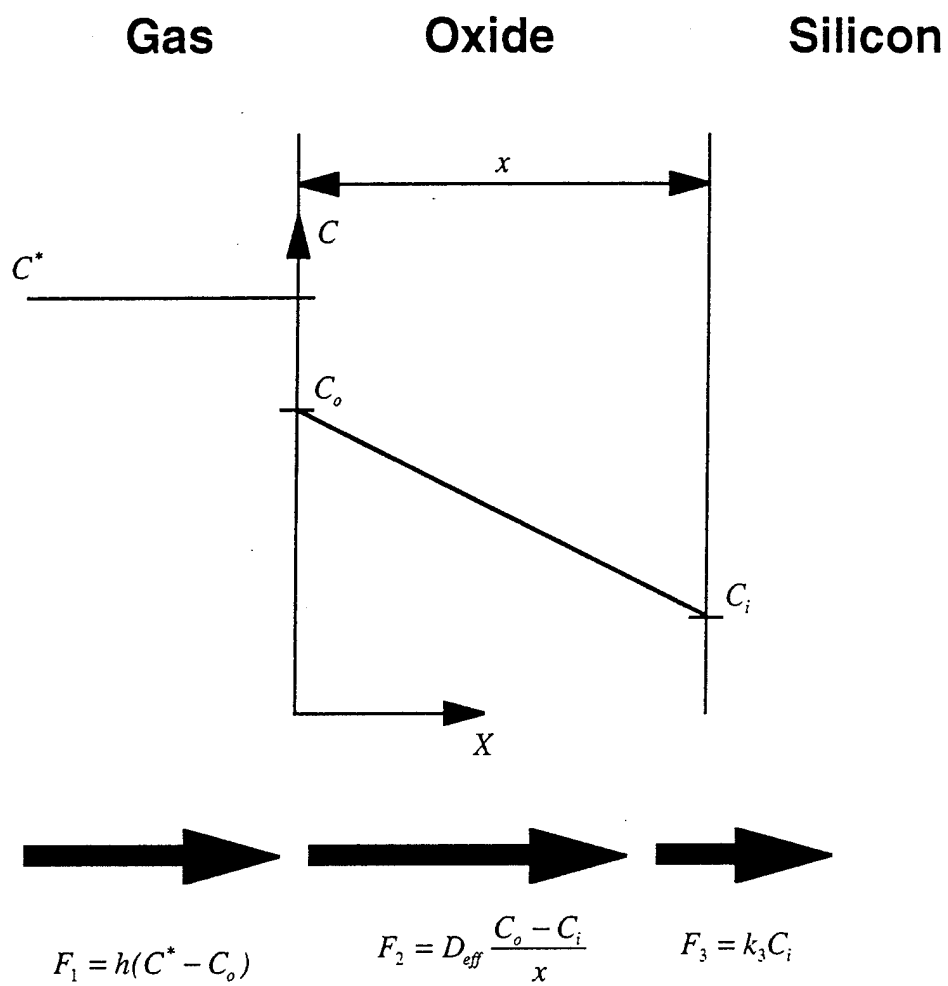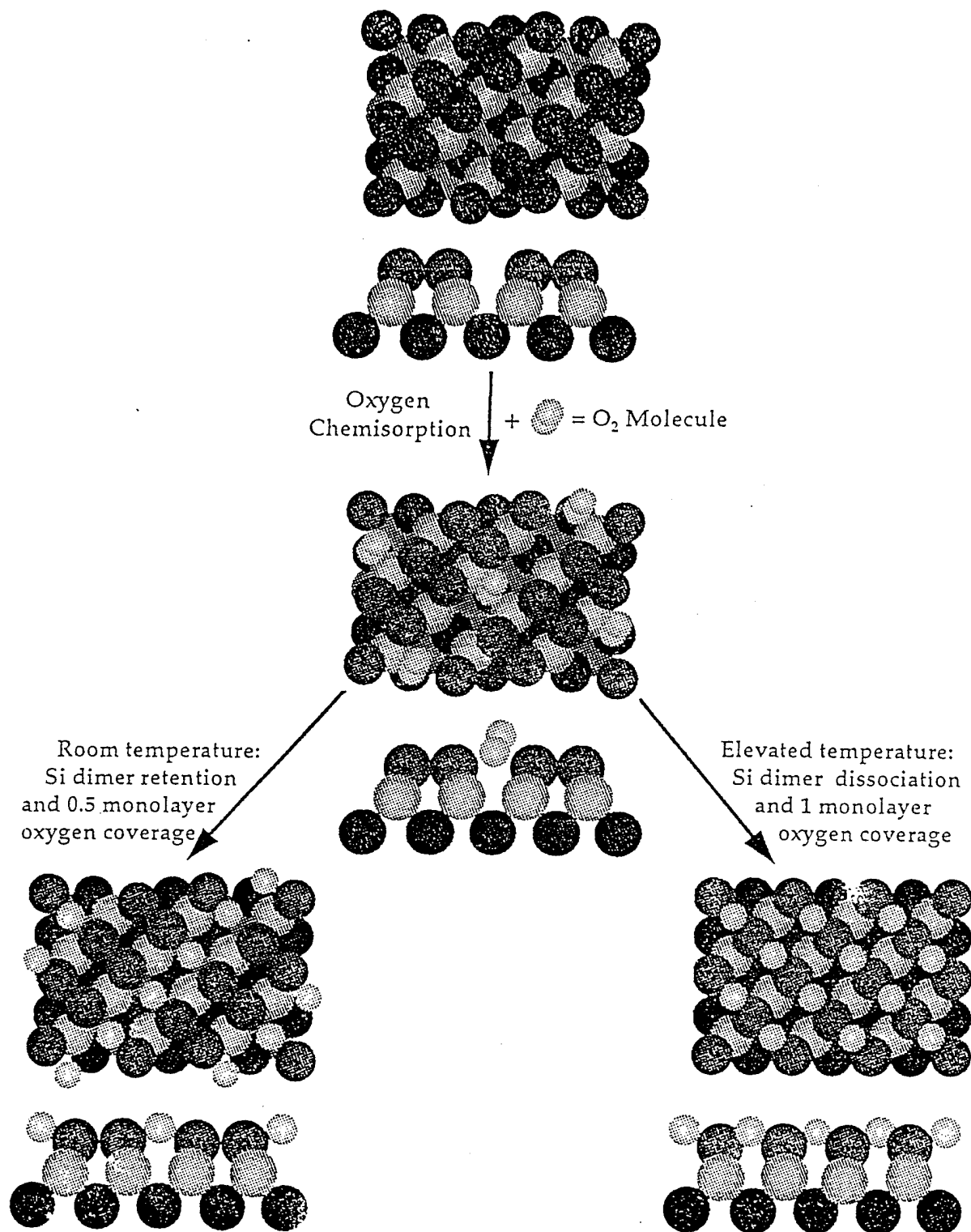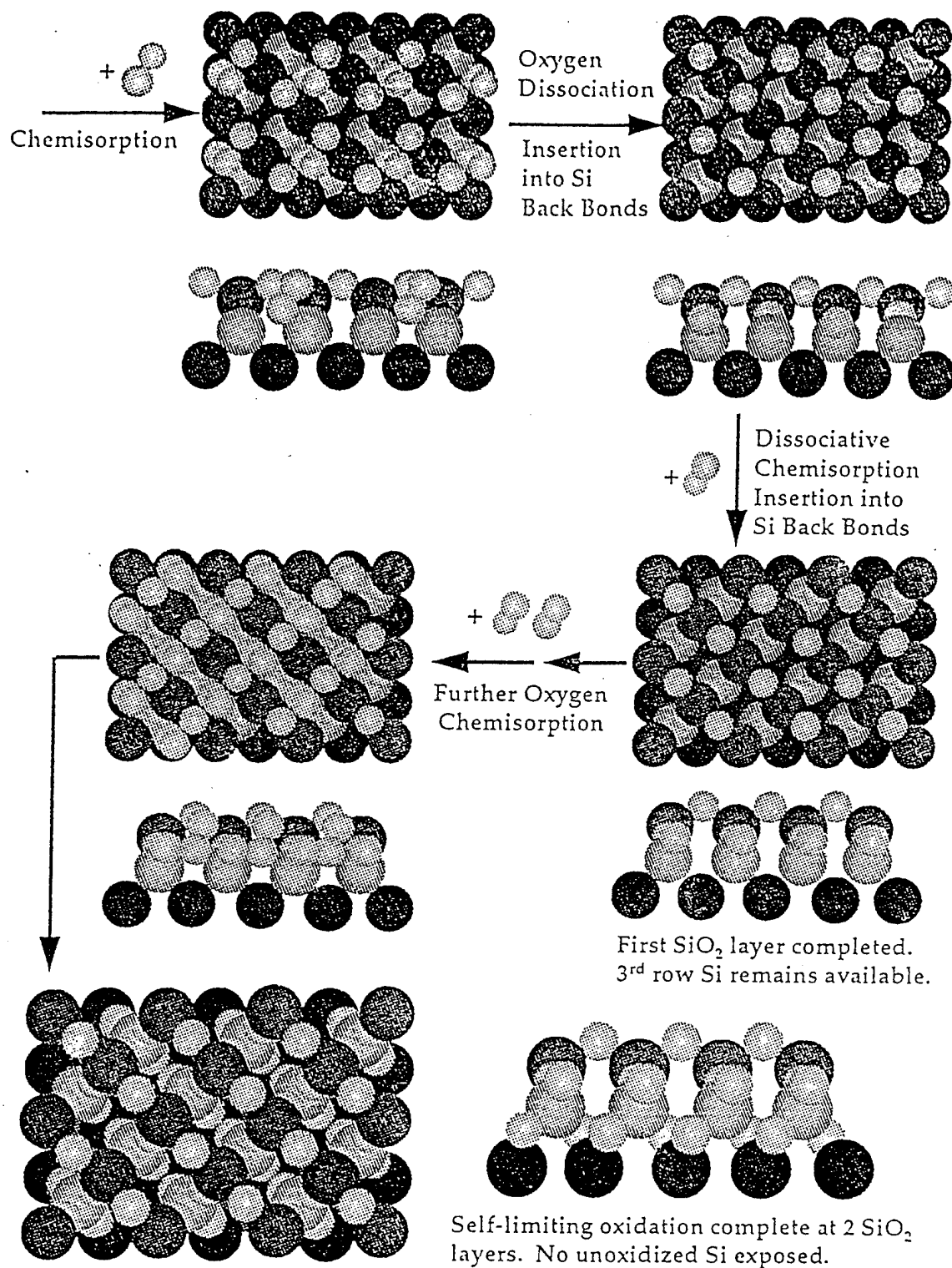
Gas   **Oxide**   **Silicon**

$$F_1 = h(C^* - C_o) \qquad F_2 = D_{\mathit{eff}} \frac{C_o - C_i}{x} \qquad F_3 = k_3 C_i$$

Figure 1. The flux picture.

Oxygen
Chemisorption + ◯ = O₂ Molecule

Room temperature:
Si dimer retention
and 0.5 monolayer
oxygen coverage

Elevated temperature:
Si dimer dissociation
and 1 monolayer
oxygen coverage

Figure 2.

Chemisorption

+ ⬤⬤

Oxygen
Dissociation

Insertion
into Si
Back Bonds

+ ⬤⬤

Dissociative
Chemisorption
Insertion into
Si Back Bonds

+ ⬤ ⬤

Further Oxygen
Chemisorption

First SiO$_2$ layer completed.
3rd row Si remains available.

Self-limiting oxidation complete at 2 SiO$_2$
layers. No unoxidized Si exposed.

Oxygen Dissociation and Insertion

Figure 3.

The Origins of Interfacial Roughness
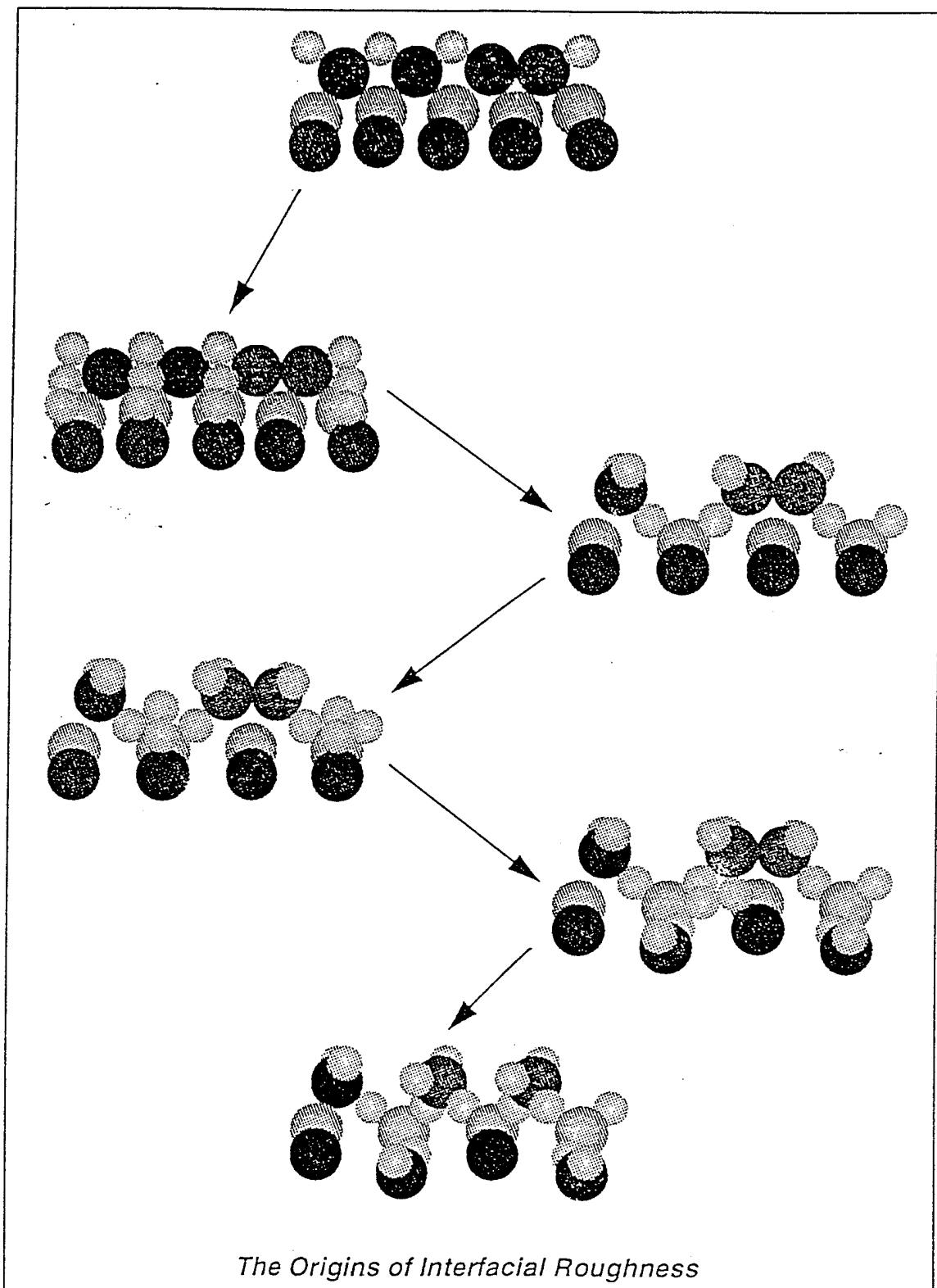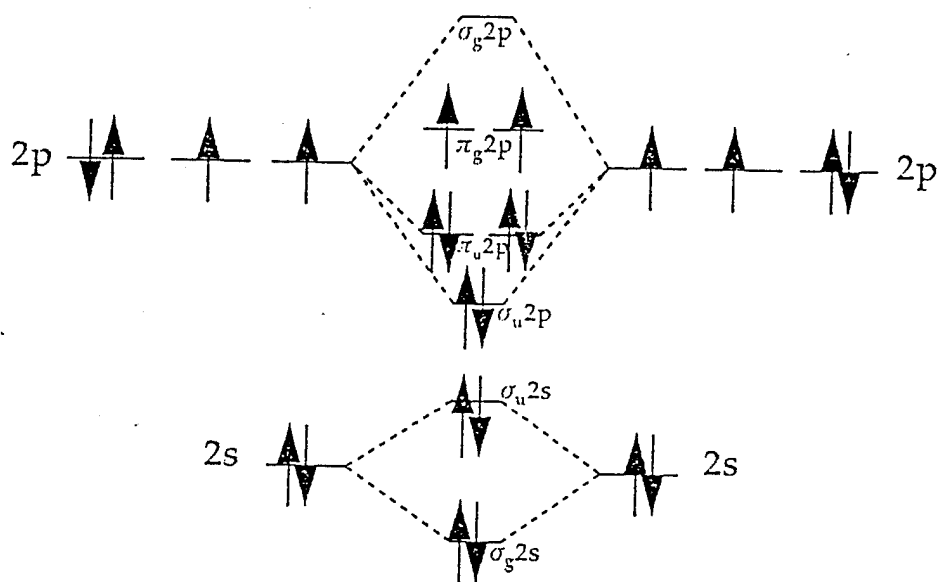
Figure 5.

*Figure 4(a) Molecular-orbital energy diagram
for the dioxygen molecule*

$\pi_g(2p_y)$

$\pi_g(2p_z)$

$\pi_u(2p_y)$
$\pi_u(2p_z)$

$\sigma_u(2p_x)$
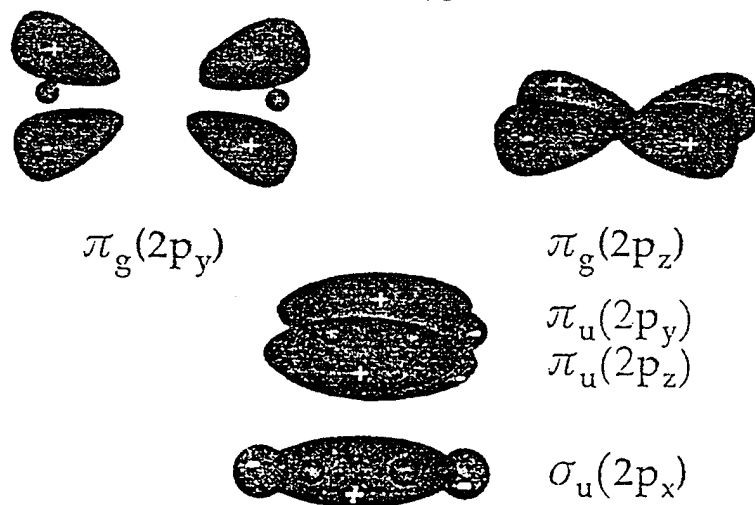
Figure 4(b) 95% probability surfaces for molecular orbitals

Figure 4.
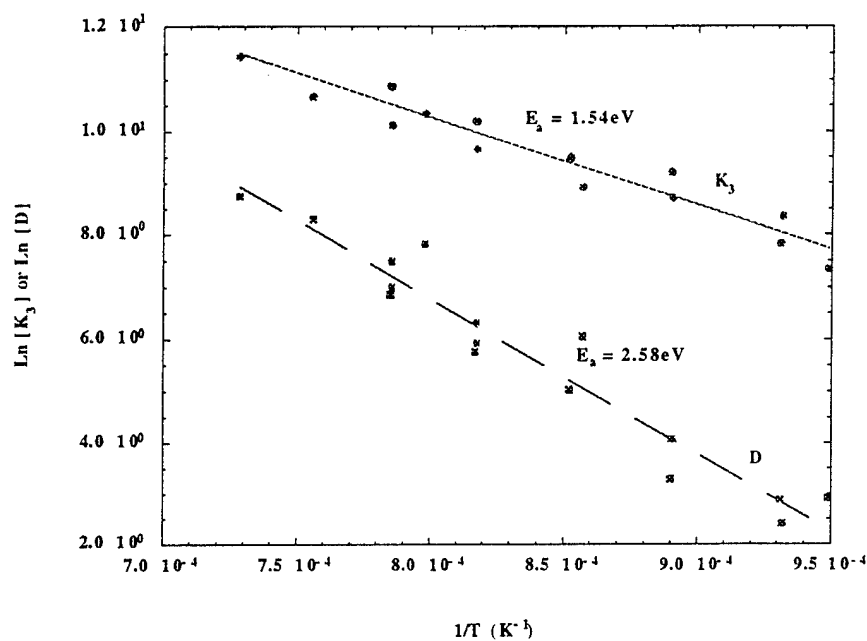
Figure 6. Activation energies [34].
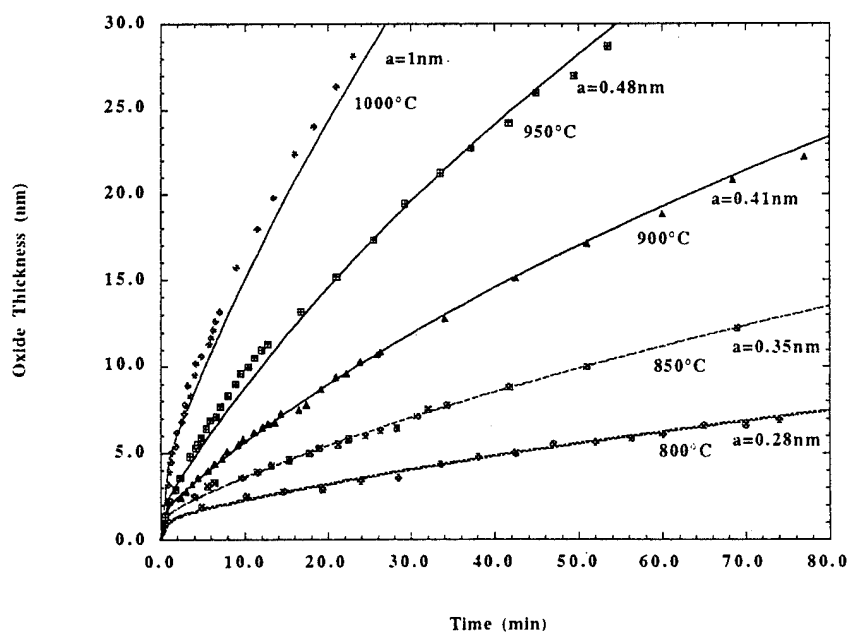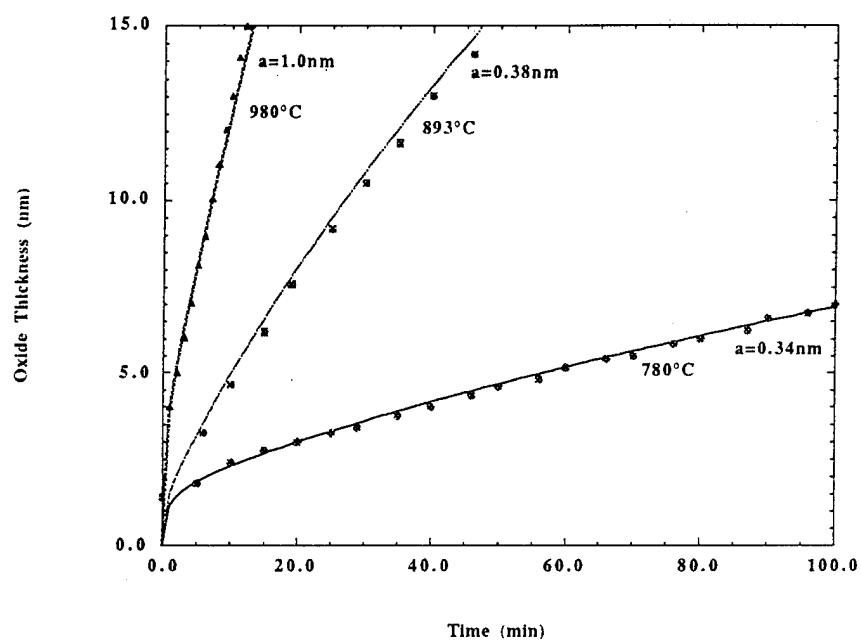

Figure 7. Model vs. Massoud's[4] data.

Figure 8. Model vs. Irene's[8] data.
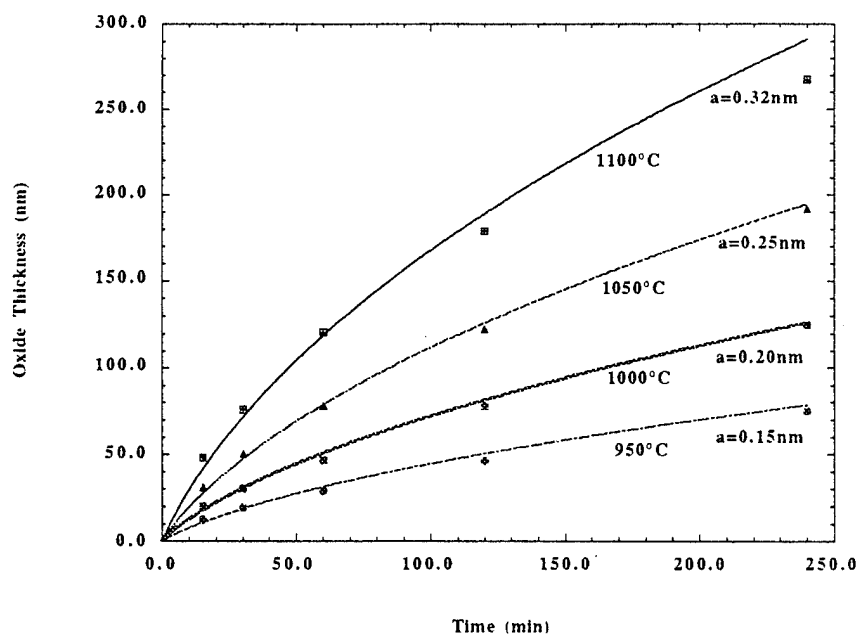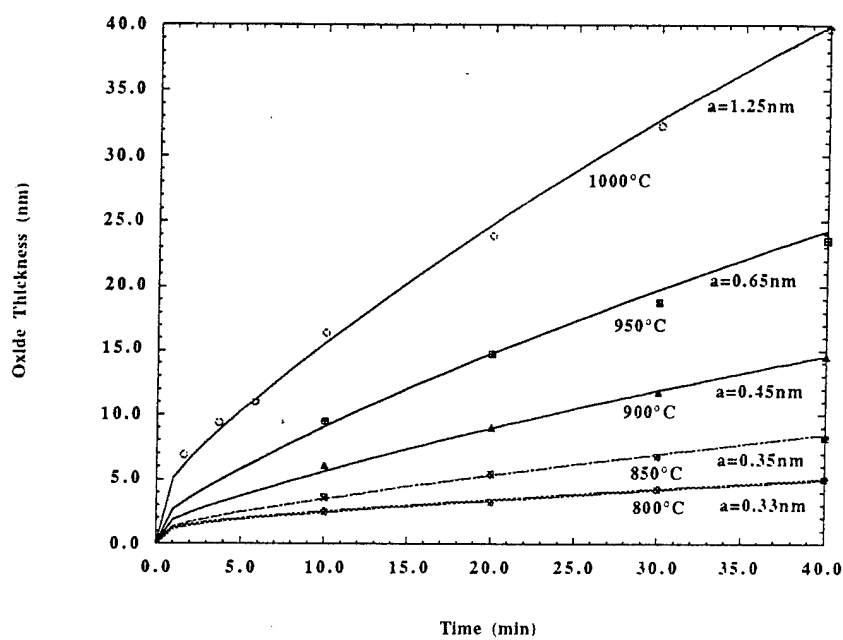

Figure 9. Model versus Kamigaki's [29] data.

Figure 10. Model versus Chao's [30] data.

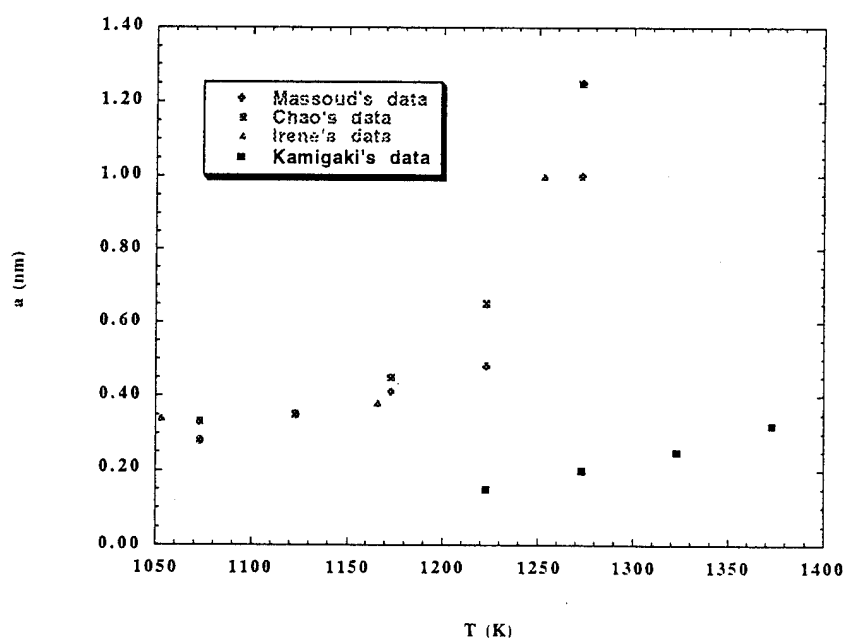**Figure 11. Dependence of *a* on temperature.**

THE COMBINATION OF EQUIPMENT SCALE AND FEATURE SCALE MODELS FOR CHEMICAL VAPOR DEPOSITION VIA A HOMOGENIZATION TECHNIQUE

Matthias K. Gobbert*, Timothy S. Cale[†], Christian A. Ringhofer[‡]

* Department of Mathematics, Arizona State University, Tempe, AZ 85287-1804, (602) 350-9250, gobbert@math.la.asu.edu, Fax (602) 965-8119

[†] Center for Solid State Electronics Research, Arizona State University, Tempe, AZ 85287-6206, (602) 965-3708, cale@enws165.eas.asu.edu, Fax (602) 965-8118

[‡] Department of Mathematics, Arizona State University, Tempe, AZ 85287-1804, (602) 965-3741, ringhofer@asu.edu, Fax (602) 965-8119

Abstract:

In the context of semiconductor manufacturing, chemical vapor deposition (CVD) denotes the deposition of a solid from gaseous species via chemical reactions on the wafer surface. In order to obtain a realistic process model, this paper proposes the introduction of an intermediate scale model on the scale of a die. Its mathematical model is a reaction-diffusion equation with associated boundary conditions including a flux condition at the micro structured surface. The surface is given in general parameterized form. A homoganization technique from asymptotic analysis is used to replace this boundary condition by a condition on the flat surface to make a numerical solution feasible. Results from a mathematical test problem are included.

Keywords:

homogenization, asymptotic analysis, finite differences, partial differential equations, chemical vapor deposition, chemical engineering.

INTRODUCTION

To model chemical vapor deposition (CVD) in single wafer reactors, Attempts have been made at linking reactor scale models (RSM) and feature scale models (FSM) to obtain realistic simulation results [1]. In these studies, reactor scale predictions have been used as inputs to feature scale models, but no information was fed back from the feature scale to the reactor scale. But features are typically arranged in clusters, which remains unaccounted for in this approach. Also, any direct combination of these models must suffer from the vast differences in length scales between the reactor scale (10^{-1}m) and the feature scale (10^{-6}m).

Therefore, we propose the introduction of a mesoscopic scale model (MSM) on the scale of a die to remedy these problems. A schematic is shown in Figure 1. To obtain an integrated process simulator, MSMs encompassing several clusters of features each are introduced at several positions on the wafer bridging the length scale differences between the reactor scale and the feature scale. By encompassing several feature clusters, a MSM also accounts for the effects of varying density of feature clustering and of cluster spacing. For the study of these feature-to-feature effects, the MSM can also be used in a stand-alone mode. *Figure 1*

Mathematically, the domain of the MSM is comprised of the gas phase just above one die on the wafer surface. Assuming that the pressure is sufficiently high, the model consists of a reaction-diffusion equation with associated boundary conditions for each chemical species in the gas domain; specifically at the wafer surface, a flux condition is imposed. It is this boundary condition that makes the problem numerically challenging, since it is impossible even for a die scale model to accurately resolve features on the scale of the micro structure.

We solved this problem using a homogenization technique from asymptotic analysis, which allows for the replacement of the micro structured surface by a flat surface by taking into account the increase in surface area inside the feature clusters. The mathematical derivation for surfaces that can be expressed in functional form has been given in [3]. This paper extends the derivation to arbitrary surfaces in parameterized form, which allows for instance for overhangs at the sides of the features.

DERIVATION

Mathematically, the problem is given as a reaction-diffusion equation with associated boundary conditions. After a suitable non-dimensionalization procedure, the dimensionless concentration $u(x, y)$ has to satisfy the differential equation

$$\frac{\partial u}{\partial t} = \text{div}_{xy}(D(x, y) \nabla_{xy} u) + R_g(u, x, y) \tag{1}$$

as well as the boundary conditions

$$-e_1^T(-D \nabla_{xy} u) = 0 \qquad x = 0, \tag{2}$$

$$e_1^T(-D \nabla_{xy} u) = 0 \qquad x = X, \tag{3}$$

$$u = c^{top}(x) \qquad y = Y, \tag{4}$$

$$\nu^T(-D \nabla_{xy} u) = S(u, x, y) \qquad (x, y) \in \Gamma_w. \tag{5}$$

D is a symmetric positive definite diffusivity matrix and $e_1 = (1, 0)^T$ the first unit vector. Γ_w denotes the parameterized wafer surface and ν the outer unit normal vector on Γ_w.

The wafer surface Γ_w is parameterized with the macroscopic variable s as

$$(x, y) = (s + \varepsilon \alpha(s/\varepsilon), \varepsilon \beta(s, s/\varepsilon)) \quad 0 \leq s \leq 1. \tag{6}$$

In a fully periodic surface, the surface would be assumed to be periodic in s with period ε, where $0 < \varepsilon \ll 1$ is a "small" quantity. However, the surface varies on both the macro scale as well as the micro scale; this fact is explicitly modeled by the dependence on the slowly changing (macroscopic) variable s and the fast changing (microscopic) variable $\sigma = s/\varepsilon$, respectively. In the definition of x, $\varepsilon \alpha(\sigma)$ represents then a microscopic perturbation of s, which allows for instance for overhangs at the feature sides on the micro scale. In y, $\varepsilon \beta(s, \sigma)$ models the microscopic surface height depending on both the macroscopic and microscopic parameterization. This parameterization allows for instance for overhangs in the surface structure. Since the surface can clearly be very different from one region to another (macroscopically), periodicity is only assumed in the fast changing variable σ (microscopically).

Hence, all surface related quantities are assumed to be periodic with period 1 in the fast changing variable, in particular the surface variables α and β satisfy

$$\alpha(\sigma + 1) = \alpha(\sigma), \quad \beta(s, \sigma + 1) = \beta(s, \sigma) \quad \text{for all } 0 \leq s \leq 1. \tag{7}$$

This really means that adjacent features are assumed to be identical, while distant features can be different. It is assumed that the parameterization is well-defined.

The idea behind the homogenization technique to be used is the elimination of dependence on the microscopic parameter σ. To this end, the surface representation is formally inflated to three dimensions depending on the two parameters s and σ independently, that is no relationship between s and σ is assumed now. This results in a three-dimensional representation of the wafer surface $\tilde{\Gamma}_w$, parameterized by s and σ independently:

$$(x, y, \xi) = (s + \varepsilon\alpha(\sigma), \varepsilon\beta(s, \sigma), \sigma + \alpha(\sigma)) \quad 0 \leq s \leq 1, 0 \leq \sigma \leq 1/\varepsilon. \tag{8}$$

For this surface, a homogenization technique like in [3] is used to find the appropriate problem for the leading term of the bulk solution $\tilde{w}_0$:

$$\frac{\partial \tilde{w}_0}{\partial t} = \text{div}_{xy}(D(x, y) \nabla_{xy} \tilde{w}_0) + R_g(\tilde{w}_0, x, y) \tag{9}$$

with the boundary conditions

$$-e_1^T(-D \nabla_{xy} \tilde{w}_0) = 0 \qquad \text{at } x = 0, \tag{10}$$

$$e_1^T(-D \nabla_{xy} \tilde{w}_0) = 0 \qquad \text{at } x = X, \tag{11}$$

$$\tilde{w}_0 = c^{top}(x) \qquad \text{at } y = Y, \tag{12}$$

$$-e_2^T (-D \nabla_{xy} \tilde{w}_0) = \tilde{\sigma}S \qquad \text{at } y = 0, \tag{13}$$

where $\tilde{\sigma}$ is given by

$$\tilde{\sigma} = \int_0^1 \sqrt{\left(1 + \frac{\partial\alpha}{\partial\sigma}\right)^2 + \left(\frac{\partial\beta}{\partial\sigma}\right)^2} \, d\sigma. \tag{14}$$

This is the simplified problem on a rectangular domain that can be efficiently solved by numerical methods. The key is that the effect of the micro structured surface is summarized into a macroscopic correction factor in the flux condition at the wafer surface.

NUMERICAL DEMONSTRATION

To demonstrate the method, a mathematical test problem has been solved with the dimensionless parameters ε chosen sufficiently large to allow for a classic solution by full resolution of the surface. First, the results denoted by the solid lines in Figures 2 and 3 are obtained from a solution of the simplified problem (9)–(13). Second, the dotted lines represent the solution obtained by solving the original problem (1)–(5) after transforming its domain onto a rectangle.

The problem uses $\varepsilon = 1/16$ and the surface function $y = \bar{h}(x) = \varepsilon h(x, x/\varepsilon), \quad h(x, \xi) = 4x(1-x)\sin(\omega_0\xi)$ with $\omega_0 = 32\pi\varepsilon$ on the unit square, hence $X = Y = 1$.

Figure 2 shows the concentration levels throughout the domain. They clearly agree everywhere except in a region close to the surface, where oscillations are introduced by actually resolving the surface structure. Figure 3 shows that the net flux into the surface predicted by the asymptotic solution captures the average of the true flux. Both facts demonstrate that the method is capable of modeling the quantities relevant to the interfaces with the reactor scale model (via the concentration levels near the top) and the feature scale model (via the average flux). *Figure 2* *Figure 3*

A demonstration of the capability of the mesoscopic scale model to study feature-to-feature effects for a physical example is contained in [2].

ACKNOWLEDGEMENTS

REFERENCES

1. T. S. Cale, J.-H. Park, T. H. Gandy, G. B. Raupp, and M. K. Jain. "Step coverage predictions using combined reactor scale and feature scale models for blanket tungsten LPCVD." *Chemical Engineering Communications*, 119:197–220, 1993.

2. M. K. Gobbert, T. S. Cale, and C. A. Ringhofer. "One approach to combining equipment scale and feature scale models." In M. Meyyappan, D. J. Economou, and S. W. Butler, editors, *Process Control, Diagnostics, and Modeling in Semiconductor Manufacturing*, pages 553–563, Reno, NV, May 1995. Electrochemical Society, 187th Meeting.

3. M. K. Gobbert and C. A. Ringhofer. "An asymptotic analysis for a model of chemical vapor deposition on a micro structured surface." Submitted to *SIAM Journal on Applied Mathematics*.

BIOGRAPHIES

Matthias K. Gobbert is a graduate student in the Department of Mathematics at Arizona State University. He is working on the numerical solution of partial differential equations arising in semiconductor process modeling.

Timothy S. Cale is a professor of Chemical Engineering and Interim Director of the Center for Solid State Electronics Research at Arizona State University. His research combines experiments with modeling and simulations to develop fundamental understanding of transport and reaction phenomena.
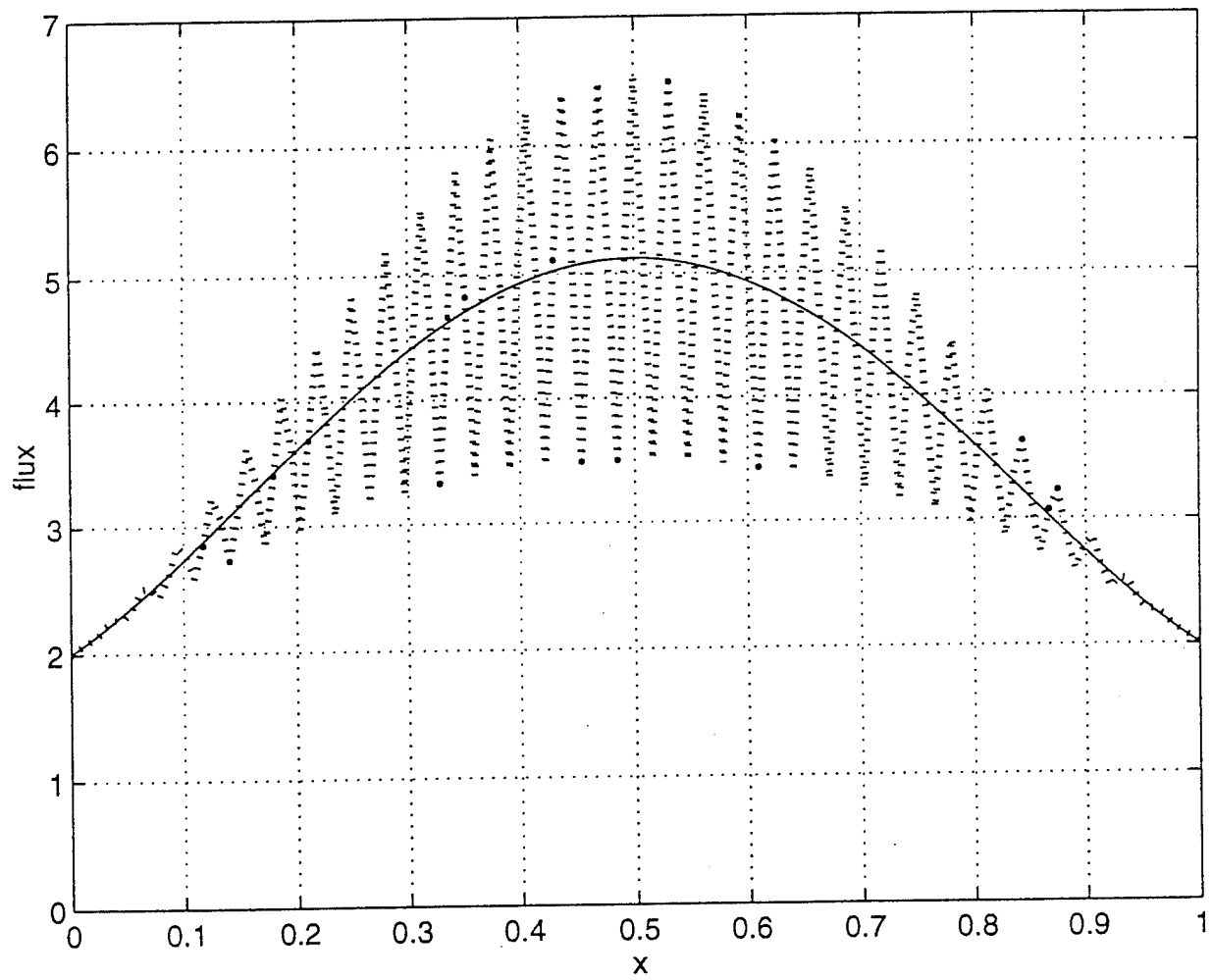
Christian A. Ringhofer is a professor of Mathematics at Arizona State University. His work includes asymptotic and numerical methods for the solution of partial differential equations.

LIST OF CAPTIONS

1. Schematic of the reactor with flow pattern and typical domains of the models.

2. Contour plot of solutions.

3. Flux plot of solutions.

RSM   MSM   FSM
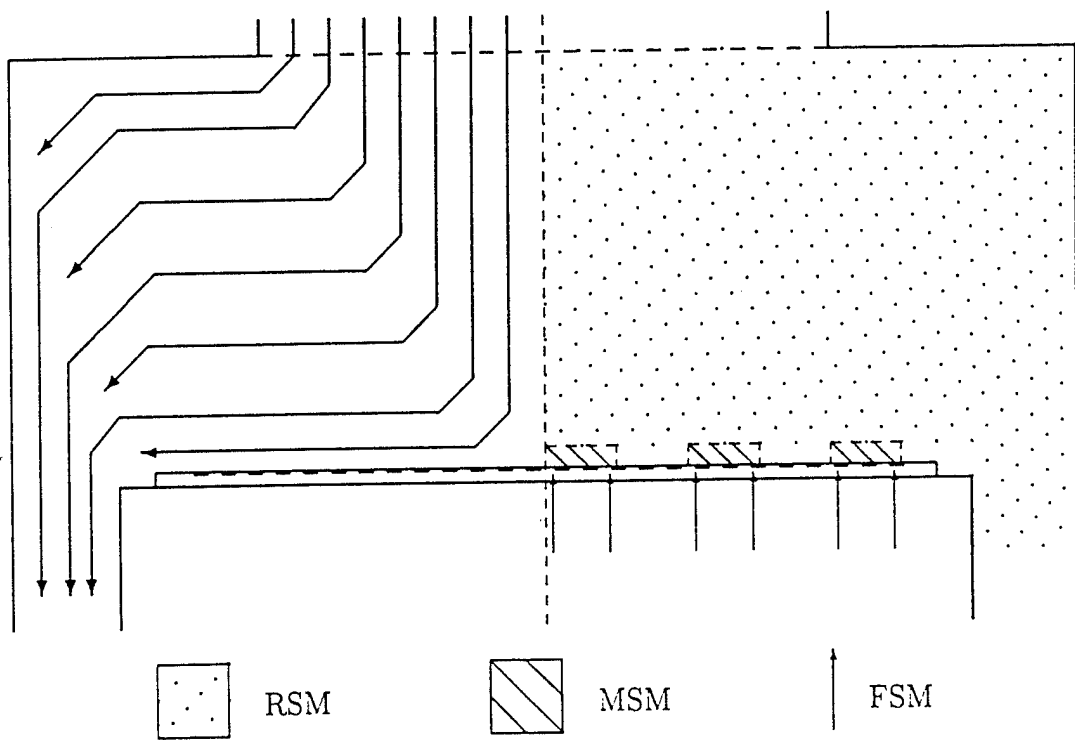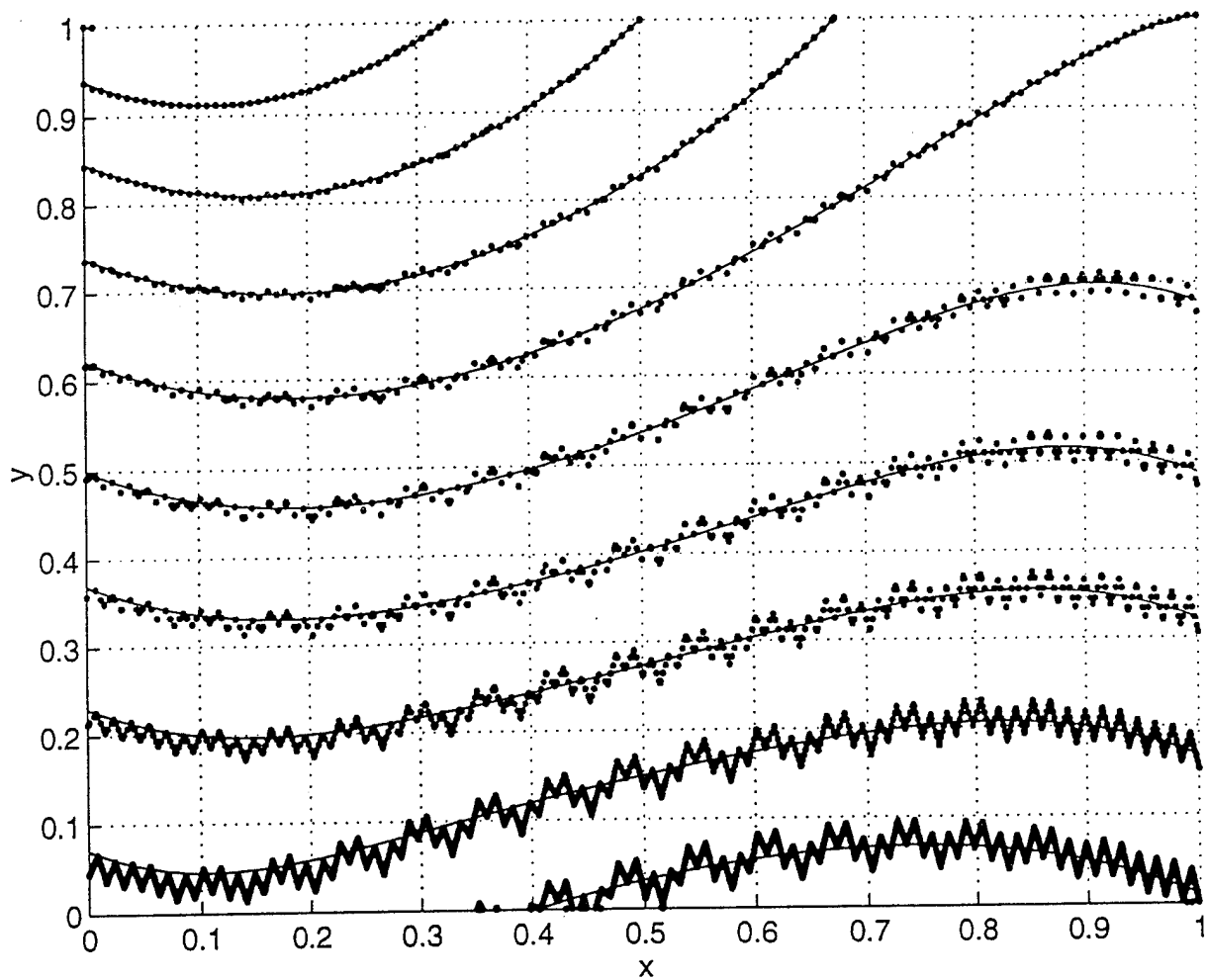
MESOSCOPIC SCALE MODELING OF MICROLOADING DURING LPCVD

Matthias K. Gobbert*, Christian A. Ringhofer

* corresponding author
Department of Mathematics, Arizona State University, Tempe, AZ 85287-1804,
phone 965-7194, e-mail gobbert@math.la.asu.edu, fax 965-8119

Timothy S. Cale

Center for Solid State Electronics Research, Arizona State University,
Tempe, AZ 85287-6206, phone 965-3708, e-mail cale@enws165.eas.asu.edu, fax 965-8118

This paper discusses a model designed to deal with pattern dependencies of deposition processes. It is a mesoscopic scale model in the sense that it deals with spatial scales on the order of 10^{-3} m to 10^{-2} m, which is intermediate between reactor scale and feature scale. This model accounts for the effects of the microscopic surface structure via suitable averages obtained by a homogenization technique from asymptotic analysis. Two studies on the LPCVD of silicon dioxide from tetraethoxysilane are presented to demonstrate the mesoscopic scale model. The first study shows the effects of microloading in regions of higher feature density. The second study shows the effects of varying operating conditions on loading and introduces a generalized Damkoehler number, which includes information about the surface patterns, to quantify the degree of transport limitations. Some thoughts on how this model can be used to bridge reactor scale and feature scale models are presented.

INTRODUCTION

The trend towards single wafer reactors (SWR) for deposition and etch processes in the microelectronics industry is expected to continue as wafer size continues to increase. In order to make single wafer reactors more economically attractive, deposition and etch processes are being run at high rates to maintain reasonable throughputs.High rates can lead to nonuniformities on the wafer scale because of depletion of reactants and transport limitations. Reactor scale models (RSM) for flow, heat transfer, and chemical reactions are well developed for single wafer reactors used for thermally driven chemical processes and susceptor based heating. In fact, reactor scale modeling and simulation have been used to help design reactors and establish operating conditions which provide acceptable wafer state uniformities for specific processes [1].

1

Although not central to the purpose of this paper, the single wafer reactor considered is assumed to be radially symmetric and in stagnation point flow. The wafer rests on a heated susceptor in the center of the reactor chamber, and the reactant gases are introduced through a shower head at the top. A schematic cross-section with a rough sketch of the flow pattern is shown in Figure 1.

High deposition rates can also lead to nonuniformities on the feature scale, even in the absence of wafer scale nonuniformities; i.e., the film thickness may not be uniform inside features. This loss of conformality is well understood and feature scale models (FSM) and simulators are fairly well developed for thermally driven deposition processes [2]. Nevertheless, the predictions of feature scale simulators depend on the species fluxes into the feature from the source volume. In general, these fluxes cannot be obtained from experiments. A complete model for a deposition process would predict local wafer state based on reactor set points.

Attempts have been made to combine reactor scale and feature scale models to obtain more realistic simulation results [3, 4]. In those studies, the models were used sequentially. Namely, species concentrations predicted by the reactor scale model are used to compute the species fluxes to the wafer surface, which are used as input for a representative feature at a particular position on the wafer. No information was fed back from the feature scale to the reactor scale, e.g. by-product concentrations were computed by the reactor scale model assuming a flat surface and then used as input for the feature scale model applied to a representative feature. This approach does not allow for the effects of feature density to be taken into account. Also, any direct combination of these models must suffer from the vast differences in length scales between the reactor scale (10^{-1}m) and the feature scale (10^{-6}m).

High rates can also lead to pattern dependencies; i.e., local deposition rates might depend upon the local pattern density. In order to account for pattern dependencies, we present a mesoscopic scale model (MSM) on the scale of millimeters. To obtain an integrated process simulator, it is envisioned that the MSM will be applied over selected areas, each encompassing several clusters of features on the wafer. For each of the MSMs, there would be several FSMs, each representing a typical feature in one of the feature clusters. See Figure 1 for an example arrangement. In this way, the MSM would bridge the length scale differences between the reactor scale and the feature scale. By encompassing several feature clusters, the MSM also accounts for the effects of varying density of feature clustering and of cluster spacing to study feature-to-feature effects. To demonstrate these capabilities, the MSM can also be used in a stand-alone mode, as in this paper.

Future papers will deal with the interchange of information between the different models. It is clear that the details of the film profile will be provided by a FSM, since it resolves the length scale of an individual feature. Since the particle flow on the mesoscopic scale is much faster than the surface growth, it is justified to treat the surface as fixed in time for the solution of the MSM. The surface would then be updated periodically, as appropriate to the time scale of the surface growth, by the FSM using flux data provided by the MSM. Therefore, since this paper presents the MSM in stand-alone mode to demonstrate its basic

2

functionality, the surface is assumed given and fixed in time. Similarly, it is assumed that the information at the gas-phase interface is supplied from a RSM.

Mathematically, the domain of the MSM is comprised of the gas phase just above one or more feature clusters on the wafer surface. Assuming that the pressure is sufficiently high, the model consists of a reaction-diffusion equation with associated boundary conditions for each chemical species in the gas domain; specifically at the wafer surface, a flux condition is imposed. It is this boundary condition that makes the problem numerically challenging, since it is unreasonable even for a mesoscopic scale model (on the scale of millimeters) to accurately resolve the patterns on the scale of the features.

We use a homogenization technique from asymptotic analysis to replace the patterned surface by a flat surface by taking into account the increase in surface area inside the feature clusters. The approach rests on the observation that patterned surfaces can be viewed as possessing two length scales, one macroscopic length scale resolving the variations from one cluster to the next and one microscopic scale resolving the changes from one feature to the next. The fundamental idea of our approach is to separate the effect of the two scales, then average over the microsopic scale while retaining the influence of macroscopic variations. The averaging is based on the mathematical assumption of periodicity in the microscopic (but not in the macroscopic) variable; this corresponds to the physical assumption that all features in one cluster are identical (on the microscopic scale), while allowing for variations between the clusters (on the macroscopic scale). In a global process simulator, this assumption means that there has to be one feature scale model corresponding to each cluster of features, which are considered identical.

This paper is divided into two mathematical and three application sections. In the next section, the mathematical derivation is sketched for surfaces given in functional form for simplicity; the full mathematical derivation has been given in [5]. The extension to parameterized surfaces (allowing overhangs on the feature sides) is contained in [6]. Following that, the numerical method used in the simulations is presented. The first of the application sections introduces the example chemistry of silicon dioxide deposition from TEOS. Afterwards, two sets of results are presented with the implementation used in stand-alone mode. First, the effect of varying feature density inside clusters for one set of operating conditions is analyzed. Second, a study of varying operating conditions for one surface example is presented.

THE MATHEMATICAL MODEL

The domain for the mathematical problem is chosen to encompass several feature clusters and to extend into the gas-phase. For the purposes of the presentation here, the species concentrations just above the wafer surface are assumed to vary slowly in space laterally, and the features are assumed to be infinite trenches. However, [5] contains a general three-dimensional derivation. The pattern on the wafer is the only rough surface and is given by

the function $y = \tilde{h}(x)$. The coordinate system is chosen with x ranging from 0 to X along the wafer surface and y ranging from 0 to Y perpendicular to the surface. Figure 2 shows an example of how four feature clusters might be introduced.

For the examples considered in this paper, the pressure is high enough (Knudsen number low enough < 0.01) that the dimensionless problem for the flow of a gaseous species in and close to the boundary layer above the surface is given by a reaction-diffusion type equation. This equation reads for one species

$$\frac{\partial c}{\partial t} = -\mathrm{div} F + R_g(c, x, y), \quad F = -D(x, y) \nabla c, \tag{1}$$

where $c(x, y)$ denotes the molar concentration and $F(x, y)$ the associated species flux, where the dependence on time t is suppressed in the notation for compactness. D is the diffusivity matrix for gaseous species in the mixture and R_g the gas-phase reaction term. On both sides of the domain, no flux conditions are used

$$-e_1 \bullet F = 0 \quad \text{at } x = 0, \tag{2}$$

$$e_1 \bullet F = 0 \quad \text{at } x = X, \tag{3}$$

where $e_1 = (1, 0)^T$ denotes the first unit vector corresponding to the x-direction. The boundary condition at the gas-phase interface is given by the known function $c^{top}(x)$, which represents a (trial) solution of the reactor scale model, as

$$c = c^{top}(x) \quad \text{at } y = Y. \tag{4}$$

Along the wafer surface, the flux is given as a function of the species generation rates on the surface, namely

$$\nu \bullet F = S(c, x, y) \quad \text{at } y = \tilde{h}(x), \tag{5}$$

where ν is the unit outward normal vector. The complete problem is then given by (1) through (5) together with the definition of the domain in Figure 2. Notice that while the appropriate resolution of the differential equation (1) does not pose any major challenges for modern computers, the proper resolution of boundary condition (5) for patterns consisting of several thousand features is still unreasonable.

For the purposes of the asymptotic analysis, the surface function is written as a function of the two variables x for the macroscopic changes and $\frac{x}{\varepsilon}$ for the microscopic changes. Here, ε is some chosen dimensionless small parameter of the problem. For instance, denote the average initial feature width by x_w, and let the space between features be x_s. Using $x_p = x_w + x_s$ for shorthand, then ε is chosen as

$$\varepsilon = \frac{x_p}{X}, \tag{6}$$

where X is the width of the domain, again. Indeed, we use $x_w = x_s = x_p/2$ for simplicity. Notice that for a small $0 < \varepsilon \ll 1$, $\xi = \frac{x}{\varepsilon}$ changes much more rapidly than x itself. With this idea, the surface function, which is itself of microscopic magnitude, can be written as

$$\tilde{h}(x) = \varepsilon \, h(x, \tfrac{x}{\varepsilon}). \tag{7}$$

4

For the mathematical analysis, this function $h(x, \xi)$ that explicitly distinguishes the effect of both length scales is assumed periodic in ξ with period 1. With the definition for ε above, this means that the patterns on the wafer surface are locally periodic with period x_p in the macroscopic variable x; this assumption is local in the sense that the definition of $h(x, \xi)$ still allows for variations resulting from the dependence on the macroscopic variable x. In physical terms, neighboring features are assumed to be identical, but clusters are allowed to differ from each other. Note that the normal vector in equation (5) computes to

$$\nu = \frac{\nu_\varepsilon}{\|\nu_\varepsilon\|_2}, \quad \nu_\varepsilon = \begin{pmatrix} \frac{\partial \bar{h}}{\partial x} \\ -1 \end{pmatrix} = \begin{pmatrix} \frac{\partial h}{\partial \xi} + \varepsilon \frac{\partial h}{\partial x} \\ -1 \end{pmatrix} \tag{8}$$

for this surface function.

A homogenization technique from asymptotic analysis is now applied to separate the two length scales using the representation

$$c(x, y) = \tilde{w}_0(x, y) + \bar{w}(x, \frac{x}{\varepsilon}, \frac{y}{\varepsilon}). \tag{9}$$

In this representation, we are interested in computing the outer solution $\tilde{w}_0$ in the bulk of the domain. It depends only on the macroscopic variables x and y. The correction term is assumed to vanish except inside the boundary layer close to the surface, that is we assume that $\bar{w}(x, \xi, \eta)$ approaches zero, as η approaches infinity, for all x and ξ. Since this function is surface related, it is also assumed to be periodic in the microsopic variable ξ with period 1, that is $\bar{w}(x, \xi + 1, \eta) = \bar{w}(x, \xi, \eta)$ for all x and η.

Letting ε go to zero shows that indeed $\bar{w}(x, \xi, \eta)$ is of order ε outside the boundary layer. Moreover, a simplified boundary condition at the wafer surface can be derived that only depends on the macroscopic variables. The full problem for the bulk solution $\tilde{w}_0$ reads then

$$\frac{\partial \tilde{w}_0}{\partial t} = -\mathrm{div} \tilde{F}_0 + R_g(\tilde{w}_0, x, y), \quad \tilde{F}_0 = -D(x, y) \nabla \tilde{w}_0 \tag{10}$$

with the boundary conditions

$$\begin{array}{llll} -e_1 \bullet \tilde{F}_0 &=& 0 & \text{at } x = 0, \\ e_1 \bullet \tilde{F}_0 &=& 0 & \text{at } x = X, \\ \tilde{w}_0 &=& c^{top}(x) & \text{at } y = Y, \\ -e_2 \bullet \tilde{F}_0 &=& \tilde{\sigma}(x) \, S(\tilde{w}_0, x, y) & \text{at } y = 0, \end{array} \tag{11}$$

where $e_2 = (0, 1)^T$ denotes the second unit vector corresponding to the y-direction and where $\tilde{\sigma}(x)$ is given for all x by

$$\tilde{\sigma}(x) = \int_0^1 \|\nu_0\|_2 \, d\xi, \quad \nu_0 = \begin{pmatrix} \frac{\partial h}{\partial \xi} \\ -1 \end{pmatrix}. \tag{12}$$

Notice that condition (11) holds now along the flat surface $y = 0$. Therefore, a numerical solution of the reaction-diffusion problem given by (10) through (12) becomes tractable, since there are no patterns that need to be resolved. All effects of the microscopic variations of the surface have been absorbed into the macroscopic function $\tilde{\sigma}(x)$.

5

THE NUMERICAL METHOD

The numerical solution of the equivalent problem given by (10) through (12) uses the conservative formulation of the associated dimensionless problem transformed to the unit square. Second order centered differences in the spatial variables are used throughout, while the implicit Euler method is used to discretize the time variable. First, a pseudo steady-state solution at the initial time is computed. The system of nonlinear equations (arising from the spatial discretization) is solved by a fixed-point iteration using a relaxation technique for improved stability whenever needed. Then, solutions at later times are computed by discretizing the transient problem. Again, the system of nonlinear equations is solved by a fixed-point iteration with relaxation. A solution is accepted if it differs from the previous iterate by less than 10^{-5} in the rms-norm. Since the surface is assumed time-independent for this paper, the pseudo steady-state solution at the initial time suffices. The numerical grid uses 100 points in the x-direction along the wafer surface and 50 points in the y-direction perpendicular to the surface.

THE PHYSICAL EXAMPLE

The example chosen for numerical demonstration is thermally induced deposition of silicon dioxide from tetraethoxysilane (TEOS) on silicon wafers with oxygen as an inert gas. Although more complex kinetic models for this deposition system have been proposed, for instance in [7], we use the model detailed by Adams and Capio [8]; this simple chemistry is chosen, because it involves a single species (TEOS) and suffices to demonstrate the method. The stoichiometry is taken to be

$$\text{TEOS} \rightarrow \text{SiO}_2 + \text{by-products.} \tag{13}$$
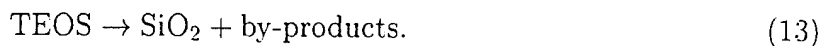
The reaction rate expression for the sole solid species SiO_2 is modeled as

$$R_{\text{SiO}_2} = k_1 \exp\left(-\frac{E_a}{RT}\right) \frac{(P_{\text{TEOS}})^{0.5}}{1 + k_2(P_{\text{TEOS}})^{0.5}} \tag{14}$$

with the coefficients

$$k_1 = 37.55 \text{ mol } / \text{ (s cm}^2\text{),} \tag{15}$$

$$k_2 = 0.25 \text{ 1 } / \sqrt{\text{torr,}} \tag{16}$$

$$E_a = 46.5 \text{ kcal } / \text{ mol,} \tag{17}$$

where E_a is the activation energy reported in by Adams and Capio [8] and where the coefficients k_1 and k_2 have been obtained by fitting a curve to the deposition rate plot in Figure 2 of [8]. Furthermore, R denotes the universal gas constant, and T is the ambient temperature. For this single-component demonstration, the net flux of TEOS to the surface is given by

$$S(c, x, t) = R_{\text{SiO}_2}(P_{\text{TEOS}}), \tag{18}$$

6

where the partial pressure of TEOS is computed from the molar concentration of TEOS via $P_{\text{TEOS}} = (c_{\text{TEOS}}/c_{total})P_{total}$. It is assumed that there are no gas-phase reactions, that is $R_g = 0$ in (1). The remaining operating conditions are the ambient temperature T and the total pressure P_{total}, both of which together determine the total concentration c_{total} via the ideal gas law. The TEOS concentration at the gas-phase interface is obtained using again the ideal gas law for a partial pressure of TEOS held constant at 0.20 torr.

The procedure used to estimate diffusivities is adopted from Reid, Prausnitz, and Poling [9, page 587] as

$$D_{ij} = \frac{\left(3.03 - \frac{0.98}{\sqrt{m_{ij}}}\right)10^{-3}T^{3/2}}{P_{total}\sqrt{m_{ij}}\sigma_{ij}^2\Omega_{D,ij}} \tag{19}$$

with the collision integral $\Omega_{D,ij}$ as suggested by Neufeld, Janzen, and Aziz [10]

$$\Omega_{D,ij} = \frac{1.06036}{(\tilde{T}_{ij})^{0.15610}} + \frac{0.19300}{\exp(0.47635\tilde{T}_{ij})} + \frac{1.03587}{\exp(1.52996\tilde{T}_{ij})} + \frac{1.76474}{\exp(3.89411\tilde{T}_{ij})}, \tag{20}$$

where

$$m_{ij} = \frac{2}{\frac{1}{m_i} + \frac{1}{m_j}}, \tag{21}$$

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2}, \tag{22}$$

$$\frac{\varepsilon_{ij}}{\kappa} = \sqrt{\frac{\varepsilon_i}{\kappa}\frac{\varepsilon_j}{\kappa}}, \tag{23}$$

$$\tilde{T}_{ij} = \frac{T}{\frac{\varepsilon_{ij}}{\kappa}}. \tag{24}$$

The variables used here are summarized in Table 1. For the approximation of the Lennard-Jones parameters σ_i and ε_i/κ following [9] as

$$\sigma_i = 1.18 \cdot (v_i)^{1/3}, \tag{25}$$

$$\frac{\varepsilon_i}{\kappa} = 1.15 \cdot T_i, \tag{26}$$

the liquid molar volume v_i and the normal boiling point T_i of species i are required. Their values for TEOS have been taken from the material safety data sheet (MSDS) and for oxygen (O_2) from appendix A (no. 59) in [9]. For convenience, these values are summarized in Table 2. Finally, the diffusivity D_{im} of species i in a multicomponent mixture is approximated as suggested by Wilke [11]

$$D_{im} = \frac{1 - x_i}{\sum_{j \neq i}\frac{x_j}{D_{ij}}}, \tag{27}$$

where the mole fraction is computed by

$$x_i = \frac{c_i}{\sum_j c_j}. \tag{28}$$

7

Then the diffusivity matrix in equation (1) is given by

$$D = \begin{pmatrix} D_{im} & 0 \\ 0 & D_{im} \end{pmatrix} \qquad (29)$$

for $i = \text{TEOS}$.

For demonstration purposes, a mesoscopic domain length of $X = 8$ mm, encompassing four feature clusters of length 1 mm each, was chosen. The domain was chosen to extend into the gas-phase to a value of $Y = 0.8$ mm. The individual features are infinite trenches of feature aspect ratio 2 with initial width 1 μm and a distance of 1 μm between features; hence, the small scale period within the feature clusters is $x_p = 2$ μm. At this spacing, there are 500 features per cluster. The dimensionless small scale parameter ε is chosen as the ratio $\varepsilon = x_p/X = 0.25 \cdot 10^{-3}$.

For the purposes of this demonstration, this structure is approximated by an explicitly given, smooth function. The dimensional form of the surface function is given by

$$y = \tilde{h}(x) = \begin{cases} a + a\, \cos\left(2\pi \frac{x}{x_p}\right) & \text{if } x \text{ is inside a feature cluster,} \\ 2a & \text{if } x \text{ is inside a flat area} \end{cases} \qquad (30)$$

with constant amplitude $a = Ax_p/4 = 1$ μm, where $A = 2$ is the feature aspect ratio of the trenches. Figure 2 shows an example of this function with a larger x_p chosen for better visibility. As an example for greater feature density, 1000 features per cluster can be approximated by replacing 2π by 4π in the argument of the cosine. Their small scale period is then $x_p = 1$ μm. Maintaining the amplitude $a = 1$ μm, their feature aspect ratio is then 4. The ratio of total area over flat area used here has been chosen for demonstration purposes, but is considered of interest to advanced manufacturing processes.

Using the definition of ε as x_p/X, a non-dimensionalization procedure with reference length X is applied to yield a definition for the function $h(x, \xi)$ with $\xi = x/\varepsilon$. From this result, the coefficient $\bar{\sigma}(x) = \int_0^1 \|\nu_0\|_2 \, d\xi$ with $\nu_0(x, \xi) = \left(\frac{\partial h}{\partial \xi}, -1\right)^T$ is computed via numerical integration using the trapezoidal rule.

THE EFFECTS OF FEATURE DENSITY IN CLUSTERS

This section demonstrates the capability of the mesoscopic scale model to study feature-to-feature effects on the scale of several feature clusters, i.e., how varying feature density inside clusters affects the concentration levels and the species fluxes at the wafer surface. The conditions are as specified in the previous section; specifically, the total pressure is 5 torr, and the temperature is 1000 K, while the partial pressure of TEOS at the top of the domain is fixed at 0.2 torr. For Case 1, the wafer surface is taken to encompass four clusters of 500 features as described in the previous section, while Case 2 uses two clusters of 1000 features and two of 500 features, each. In both cases, the steady-state solution for these

time-independent surfaces are computed by the method detailed above. Preliminary results have been shown in [12]; however, an unrealistically large value for $\tilde{\sigma}$ was used in that work.

Figures 3 and 4 show the concentration profiles throughout the two-dimensional domain. As expected, the concentration levels are lower closer to the wafer surface than at the gas-phase interface at the top. They are also lower inside the feature clusters than in the flat areas in between them; this is explained by the increase in surface reactions due to the increase in surface area available for reactions inside the feature clusters. It can be observed in Figure 4 that the concentration levels in the denser clusters are lower than in the others. However, the concentration levels in the right-hand side of the domain (encompassing the clusters with 500 features) are equal in both cases; this can be verified by the contour plots for both solutions in Figure 5, where the solid lines correspond to Case 1 and the dashed lines to Case 2.

To allow a more quantitative assessment of the solution properties, the species fluxes of TEOS into the wafer surface are shown in Figure 6. This is one of the most important quantities for the purpose of interfacing with a feature scale model as well as to characterize the quality of the solution method. Again, the solid line corresponds to Case 1 and the dashed line to Case 2. The graphs show the net fluxes *per flat wafer area*; these are the fluxes that are seen by the macroscopic scale. The right-hand side of the boundary condition on the wafer surface $\tilde{\sigma}(x)\, S(c, x, 0)$ in (11) is plotted versus the macroscopic position along the wafer surface x. As a basic observation, the flux of the depleted species TEOS into the surface is higher inside the feature clusters than in the flat areas; the transition is discontinuous, since the transition from flat area to the clusters is abrupt as well. This effect is commonly known as microloading. Quantitatively, the flux inside the feature clusters is larger by a factor of about 2.3 than in the flat areas, corresponding to the value of $\tilde{\sigma}$ there. Observe that the flux decreases in the interior of each individual feature cluster as compared to the edges of the cluster. This reflects the fact that in the center of a cluster, the least amount of TEOS is available for reaction and resulting deposition as compared to the edges, where more TEOS is available by diffusion from the flat areas; the transition between these areas is continuous, since the concentration profile changes smoothly. Finally, Figure 6 also shows that the flux of Case 2 is approximately twice as high in the clusters with 1000 features than in the others. This corresponds again to the higher value of the factor $\tilde{\sigma}$ being 4.2 in Case 2 as compared to 2.3 in Case 1, each inside the feature clusters.

In summary, the effect of the feature density increase in some of the feature clusters increases the depletion of reactants in those areas due to an increase in available surface area. This shows that the method predicts the concentration levels and species fluxes and accounts for the variations from one feature cluster to the next on the macroscopic scale. The method is also capable of representing the variations inside a feature cluster. The scale of these variations is clearly still far above the microscopic scale of an individual feature, but is one order of magnitude below that of the macroscopic scale of the overall die; note that predictions on the scale of several hundred features are not reasonably obtainable using a feature scale model.

As another application of the mesoscopic scale model, this section presents a study of the effects of varying operating conditions in the reactor. Two parameters are varied throughout a slightly larger operating window than presented in [8]. The total pressure throughout the reactor is varied from 1 torr to 9 torr, and the temperature is varied from 900 K to 1100 K. The partial pressure of TEOS at the top is again fixed at 0.2 torr as are all other parameters. In Figures 7 through 11, results for the surface of Case 1 (four clusters with 500 features each) are marked by a solid line with circles at the data points, while Case 2 (two clusters with 1000 features, two with 500 features) is marked by dashed lines with crosses at the data points; in addition to these two cases, a Case 0 representing results for a flat wafer is shown by a dotted line with stars at the data points. The results in Figure 8 are shown for Case 1, only; proportional results have been observed for the other cases. Note that the operating conditions used in the previous section lie at the center of the operating window considered here.

Figure 7 shows the average deposition rate *per total area* versus total pressure with the temperature parameterized; this is the actual depostion rate observed on the surface of the wafer. The plot shows that for all cases the deposition rate grows as the temperature increases, which is clear from the rate expression (14). It shows also that the deposition rate decreases as the total pressure increases. This can be explained by the observation that the diffusivity is inversely proportional to the total pressure, see equation (19); hence, the transport of the reactant species to the surface becomes limiting, as the total pressure increases. Moreover, the deposition rate is higher for flat wafers (dotted line for Case 0) than for patterned wafers and lower for higher density patterns (dashed line for Case 2) than for lower density patterns (solid line for Case 1).

Figure 8 plots the fractional difference in the flux versus total pressure with the temperature parameterized. This difference is defined as

$$F_{diff} = \frac{F_{max} - F_{min}}{F_{min}}, \tag{31}$$

where $F_{max} = \max_x(\tilde{\sigma}(x)\, S(c, x, 0))$, and F_{min} is defined analogously. This means that F_{min} really measures the flux level at the flat areas of the surface (see Figure 6 for instance), hence, $F_{max} - F_{min}$ measures the absolute increase in flux inside the feature clusters compared to the one associated with the flat areas. F_{diff} is then a measure for the size of the relative loading increase under the given operating conditions. It is expected that the concentration gradients will increase with increasing temperature, since the reactivity increases faster than the diffusivity. With increasing total pressure, we also expect transport limitations to restrict reactant diffusion and increase the size of the gradients. In both cases, the measure F_{diff} will decrease, since the flux to the areas with higher feature density will decrease faster than the flux to the flat areas.

Figure 9 shows the Damkoehler number Da versus total pressure with the temperature parameterized. Its basic purpose is to measure the influence of transport limitations on the

deposition process [13]. The classical definition arises in the non-dimensionalization of the flux boundary condition at the wafer surface as $\mathrm{Da} = (R_{ref}\ell_{ref})/(D_{ref}c_{ref})$; in the context of feature scale models, a similar quantity is known as the step coverage modulus [14, 15]. This definition can be used for a model on any length scale, since no information about the surface structure is used. More generally, the purpose of the Damkoehler number is to measure the characteristic deposition rate versus the characteristic transport rate. We propose therefore to take into account the information available about the surface on the mesoscopic scale under consideration. In this spirit, we define the following generalized Damkoehler number for general three-dimensional surfaces

$$\mathrm{Da} = \frac{A_D R_{ref} \ell_{ref}}{A_T D_{ref} c_{ref}}, \tag{32}$$

where A_D denotes the characteristic deposition area and A_T the characteristic transport area. A_D is then the total surface area available for deposition, i.e., the true surface area of the patterned surface, while A_T is the area that determines the transport limitations, i.e., the flat surface area. All other quantities are reference quantities arising in the non-dimensionalization procedure as in the classical definition. For the two-dimensional representation of the surface used here (infinite trenches), the ratio A_D/A_T is equal to $\tilde{\sigma}_{total} = \int_0^X \tilde{\sigma}(x)\,dx$, since this measures the total increase in surface area due to the patterns. Hence, the following formula gives the generalized Damkoehler number used

$$\mathrm{Da} = \frac{\tilde{\sigma}_{total} R_{ref} \ell_{ref}}{c_{ref} D_{ref}}. \tag{33}$$

For low temperatures and pressures, we expect the deposition to be reaction rate limited; i.e. the Damkoehler number should be low. On the other hand, for high temperatures and pressures, the Damkoehler number is expected to be high to indicate transport limitations. Both effects are exhibited by the Damkoehler number in Figure 9.

Furthermore, comparing the different surface patterns, we note that the cross-over to the transport limiting conditions ($\mathrm{Da} \approx 1$) occurs earlier for more patterned wafer surfaces. This is due to the more difficult transport into the feature clusters in these cases. This is an important consideration when comparing models on the scale of feature clusters, and the classical definition of the Damkoehler number does not capture this effect, since its value would be the same for all surfaces.

The increase in temperature affects the reaction rate term in the definition of Da in an exponential way, while the diffusivity contains the term $T^{3/2}$; the influence of the exponential function is dominant for the values of the temperature considered here, thus the exponential increase with respect to temperature seen in Figure 9. On the other hand, the change in *total* pressure only impacts the diffusivity, since all other terms (including c_{ref}) depend only on the *partial* pressure of TEOS. The diffusivity is inversely proportional to the total pressure (see equation (19)), hence the Damkoehler number goes linearly with total pressure as shown in Figure 9.

Figure 10 shows a plot of the effectiveness factor η versus the Damkoehler number Da. η is defined following [13] as the ratio of the observed deposition rate over the deposition rate

11

in the absence of gradients. The latter is taken from the reaction rate model as a function of the reference quantities, only, i.e., without considering the geometry of the surface. Thus, η is another measure of how much transport limitations affect the deposition process. It is desirable to have this measure be roughly independent of the specific surface structure. In other words, we would like to have just one curve in this graph encompassing all surface structures. Notice that this goal is approached using the generalized Damkoehler number as defined in this paper; if the classical Damkoehler number were used, the plot would split up into three branches at a lower value of Da.

An Arrhenius plot is included in Figure 11. It reflects the exponential model used for the reaction rate, since it is linearly decreasing for low pressures for all temperatures and for all pressures, if the temperature is sufficiently low. For higher pressures and high temperatures, the reaction rate per total area is again seen to be transport limited. Again, we observe that the deposition rate per total area is higher for the flat wafer and lower for the most highly patterned case compared to the case with four clusters of 500 features each. Note that for the reaction controlled regime at low temperatures, these differences become insignificant, thus demonstrating that transport limitations play a crucial role in the high temperature cases (see also Figure 7).

In summary, this study of a representative operating window demonstrates the capabilities of the mesoscopic model. This section highlights the general reason for using simulations in that any quantity can be analyzed, whether it would be actually accessible in a real reactor or not. Also, operating conditions could be chosen outside the normal, safe operating range of the reactor; however, it became clear in this project that also simulations will suffer in such cases, since the underlying physical estimation procedures can become unreliable if used outside their range of validity and better procedures are rarely available.

CONCLUSIONS

A mesoscopic scale model on the scale of several feature clusters has been introduced. Its mathematical background has been highlighted. A physical example chemistry has then been used to demonstrate the power of the model for analyzing feature-to-feature effects as well as to study an extensive window of operating conditions. In its stand-alone mode, the model has been shown to yield appropriate results.

A generalized Damkoehler number has been introduced. It was demonstrated that this number characterizes the influence of transport limitations appropriately while using available information about the surface patterns.

This paper demonstrates the principal capabilities of the method with its implementation used in stand-alone mode for a single-species chemistry. It is planned to link the implementation with existing reactor scale and feature scale codes. We also plan to demonstrate the capabilities of the stand-alone implementation for multiomponent chemistries, as well as for etch processes.

ACKNOWLEDGMENTS

REFERENCES

1. D. W. Studiner, J. T. Hillman, R. Arora, and R. F. Foster, in *Advanced Metallization for ULSI Applications 1992*, T. S. Cale and F. Pintchovski, Editors, p. 211, Materials Research Corporation, Phoenix, AZ (1993).

2. T. S. Cale and G. B. Raupp, *Vac. Sci. Tech.*, **B 8**, 1242 (1990).

3. T. S. Cale, J.-H. Park, T. H. Gandy, G. B. Raupp, and M. K. Jain, *Chemical Engineering Communications*, **119**, 197 (1993).

4. A. Hasper, J. Holleman, J. Middelhoek, C. R. Kleijn, and C. J. Hoogendoorn, *J. Electrochem. Soc.*, **138**, 1728 (1991).

5. Matthias K. Gobbert and Christian A. Ringhofer, Submitted to *SIAM Applied Mathematics*.

6. Matthias K. Gobbert, Timothy S. Cale, and Christian A. Ringhofer, Proceedings of the *Fourth International Workshop on Computational Electronics*, Tempe, AZ, November 1995.

7. Gregory B. Raupp, Frank A. Shemansky, and Timothy S. Cale, *Vac. Sci. Tech.*, **B 10**, 2422 (1992).

8. A. C. Adams and C. D. Capio, *J. Electrochem. Soc.*, **126**, 1042 (1979).

9. Robert C. Reid, John M. Prausnitz, and Bruce E. Poling, McGraw-Hill, fourth edition (1987).

10. Philip D. Neufeld, A. R. Janzen, and R. A. Aziz, *J. Chemical Physics*, **57**, 1100 (1972).

11. C. R. Wilke. *Chem. Eng. Prog.*, **46**, 95 (1950).

12. Matthias K. Gobbert, Timothy S. Cale, and Christian A. Ringhofer, in *Process Control, Diagnostics, and Modeling in Semiconductor Manufacturing*, M. Meyyappan, D. J. Economou, and S. W. Butler, Editors, **PV 95-2**, p. 553, The Electrochemical Society Proceedings Series, Pennington, NJ (1995).

13. James J. Carberry, *Chemical and Catalytic Reaction Engineering*, McGraw-Hill, New York (1976).

14. T. S. Cale, M. K. Jain, and G. B. Raupp, *J. Electrochem. Soc.*, **137**, 1526 (1990).

15. Hung Liao and Timothy S. Cale, *Vac. Sci. Tech.*, **A 12**, 1020 (1994).

List of Tables

List of Figures

Symbol	name	units
D_{ij}	binary diffusivity coefficient	cm^2 / s
T	temperature	K
P_{total}	total pressure	torr
$\Omega_{D,ij}$	collision integral	1
$\tilde{T}_{ij}$	reduced temperature	1
σ_i	Lennard-Jones parameter of species i	Å
$\frac{\varepsilon}{\kappa}$	Lennard-Jones parameter of species i	K
m_i	molecular weight of species i	g / mol
v_i	molar volume of species i	cm^3 / mol
T_i	normal boiling point of species i	K
ρ_i	mass density of species i	g / cm^3

Table 1: Variables used in the diffusivity estimation following Reid [9].

Quantity [units]	TEOS	Oxygen
molar weight m_i [g / mol]	208.3083	31.999
liquid molar volume v_i [cm^3 / mol]	222.7896	27.8494
normal boiling point T_i [K]	441	90.2
mass density ρ_i [g / cm^3]	0.935	1.149
Lennard-Jones parameter σ_i [Å]	7.1534	3.5767
Lennard-Jones parameter $\frac{\varepsilon_i}{k}$ [K]	507.15	103.73

Table 2: Material constants for the TEOS-oxygen system. The normal boiling point for both species is taken at 1 atm.
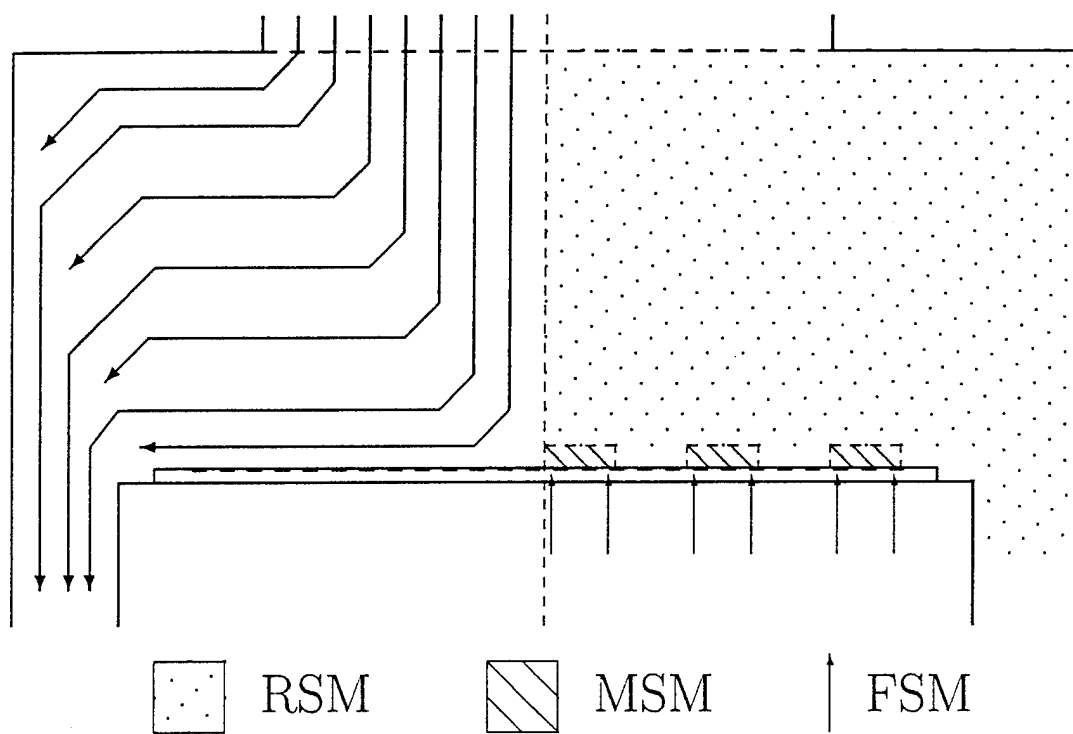
Figure 1: Schematic of the reactor with flow pattern and typical domains of the models.
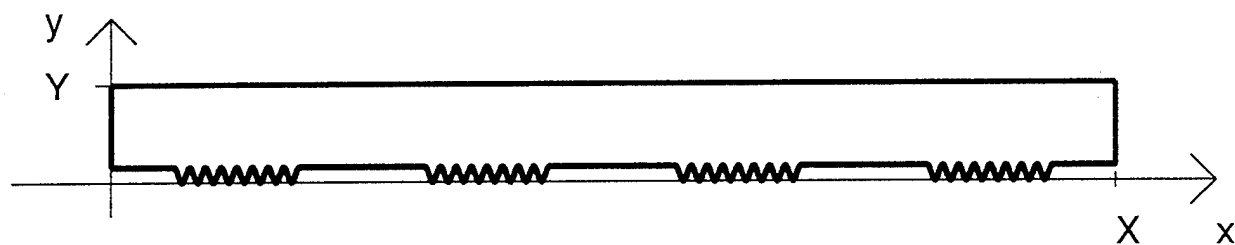
Figure 2: Sketch of a possible domain of the problem with associated coordinate system.
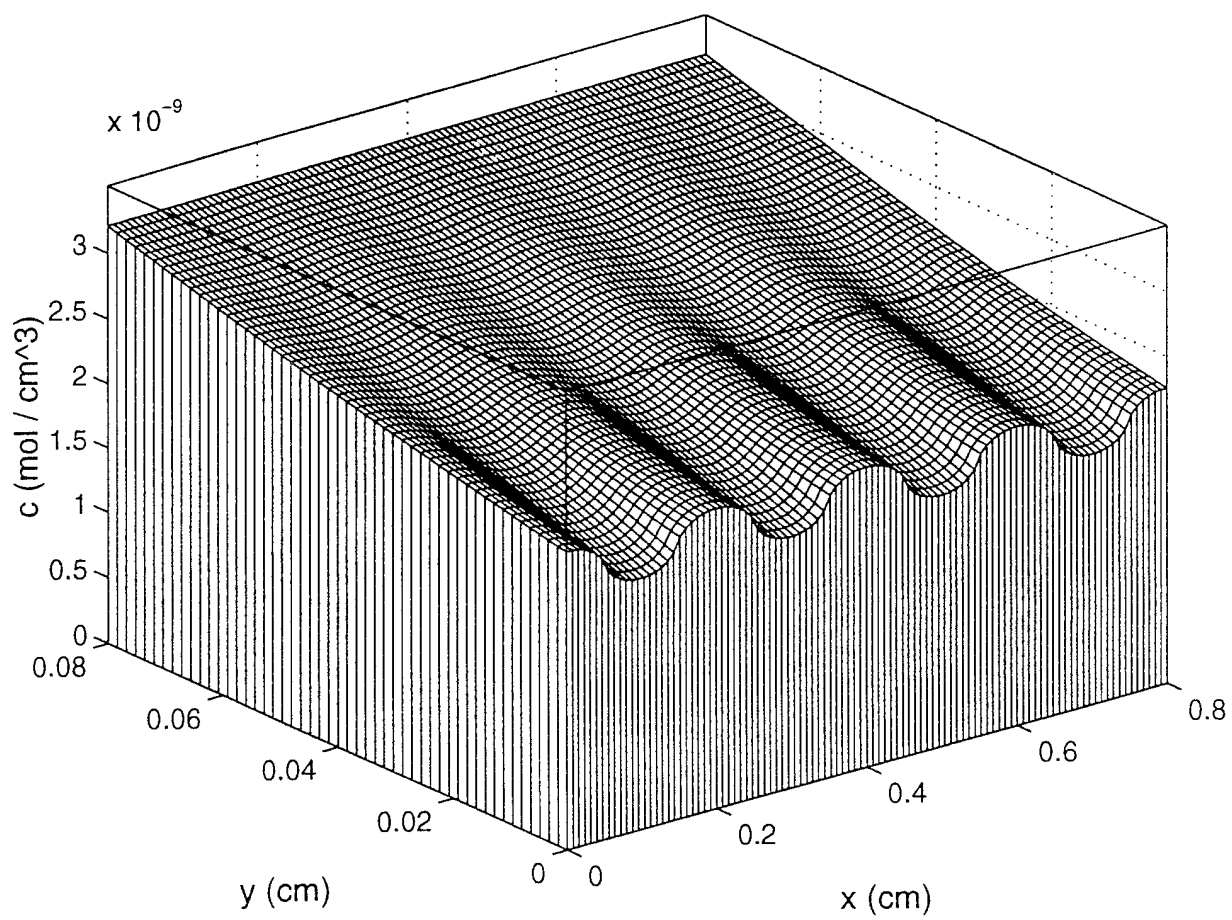
Figure 3: Concentration profile for four identical feature clusters with 500 features each.
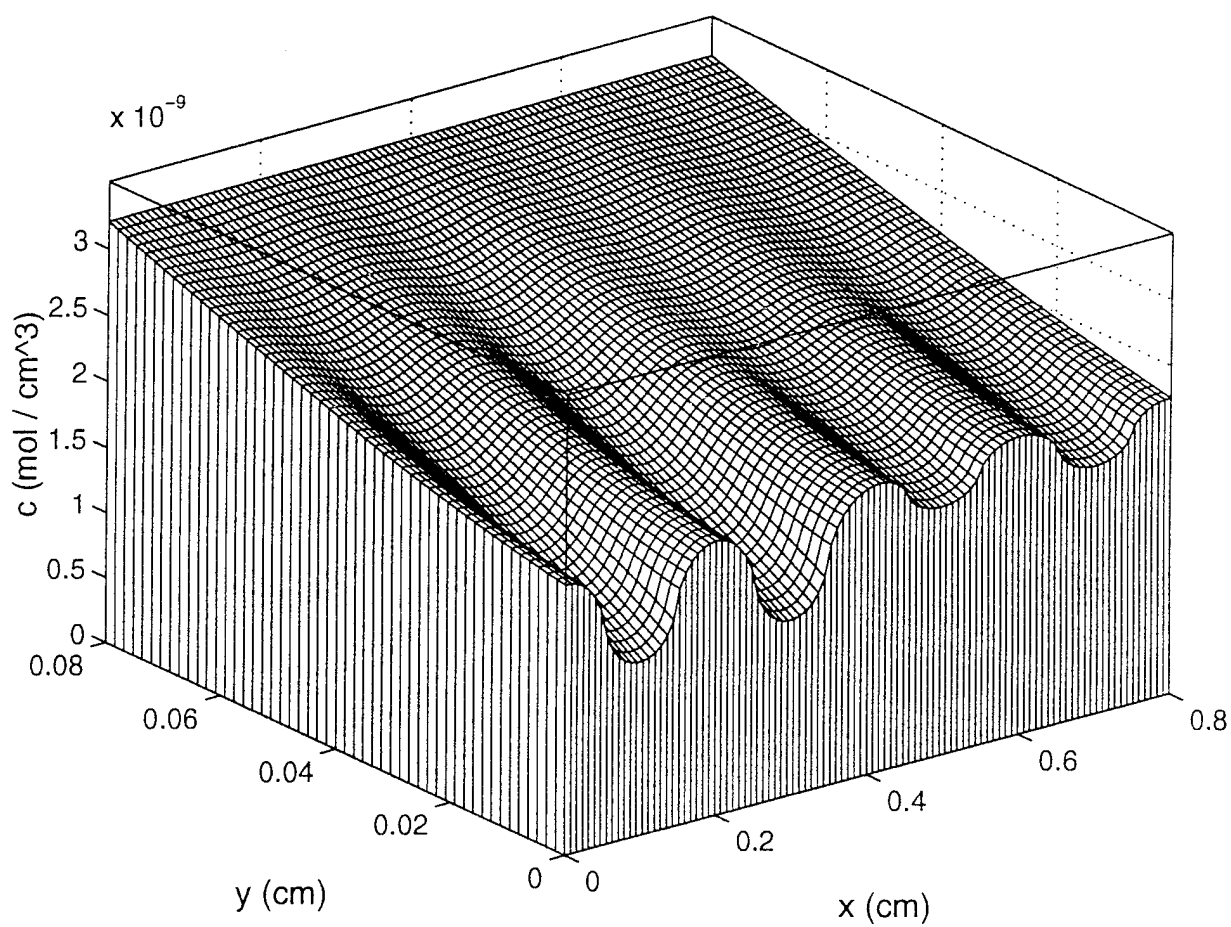
Figure 4: Concentration profile for four differing feature clusters, two with 1000 features and two with 500 features.
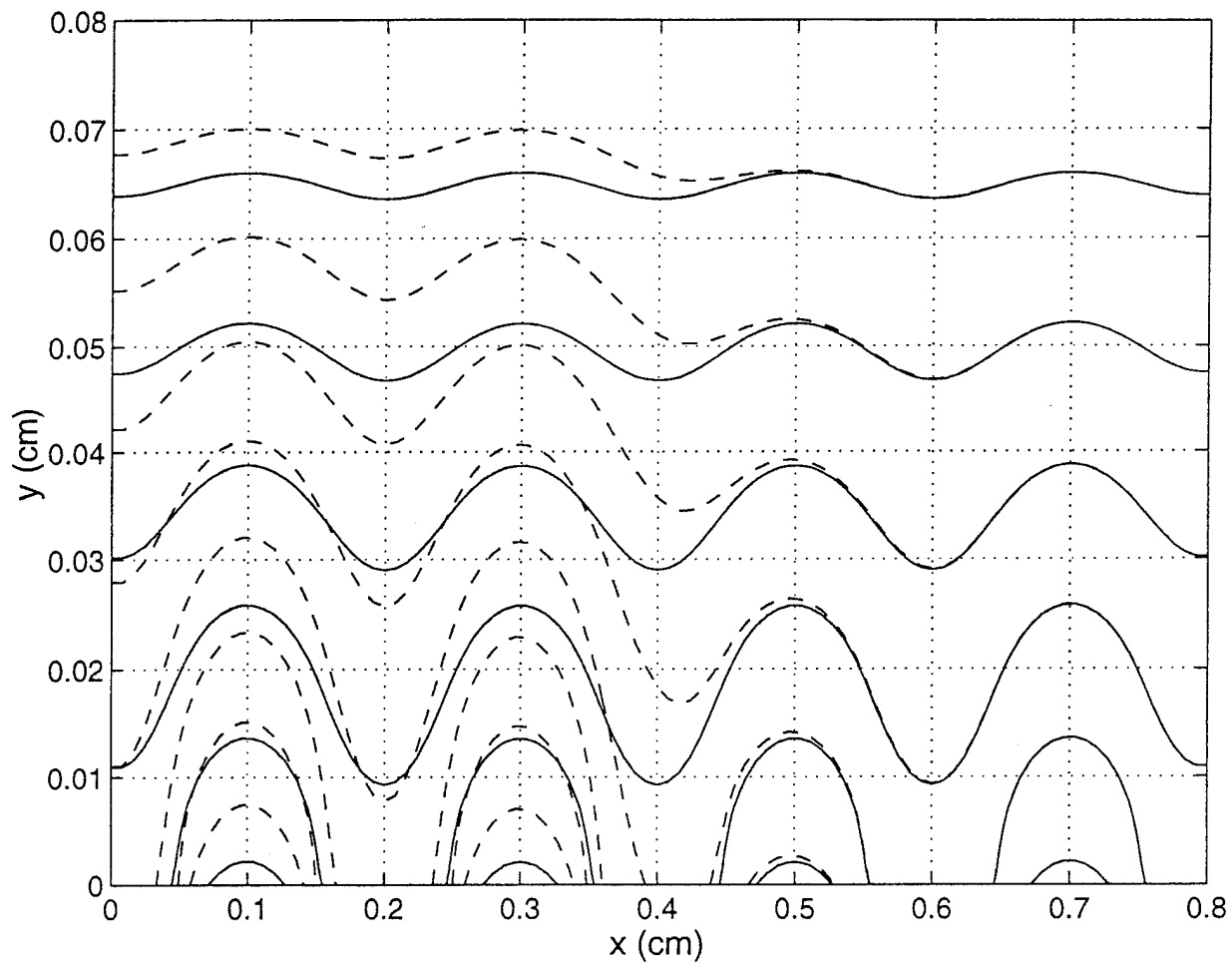
Figure 5: Comparison of concentration profiles. The solid lines depict the results for four feature clusters with 500 features each. The dashed lines mark the results for two clusters with 1000 features and two with 500 features.
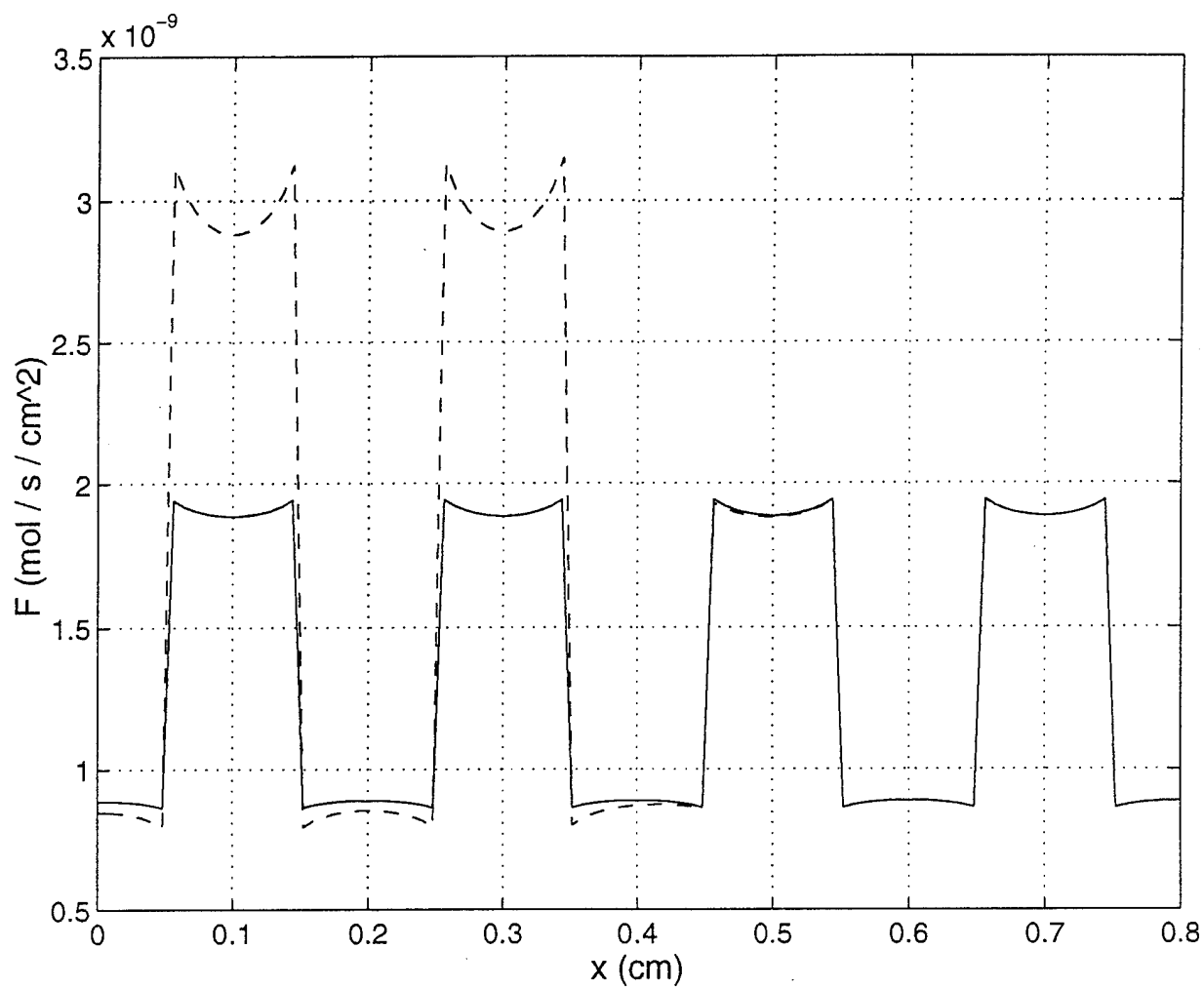
Figure 6: Comparison of net fluxes per flat area. The solid lines depict the results for four feature clusters with 500 features each. The dashed lines mark the results for two clusters with 1000 features and two with 500 features.
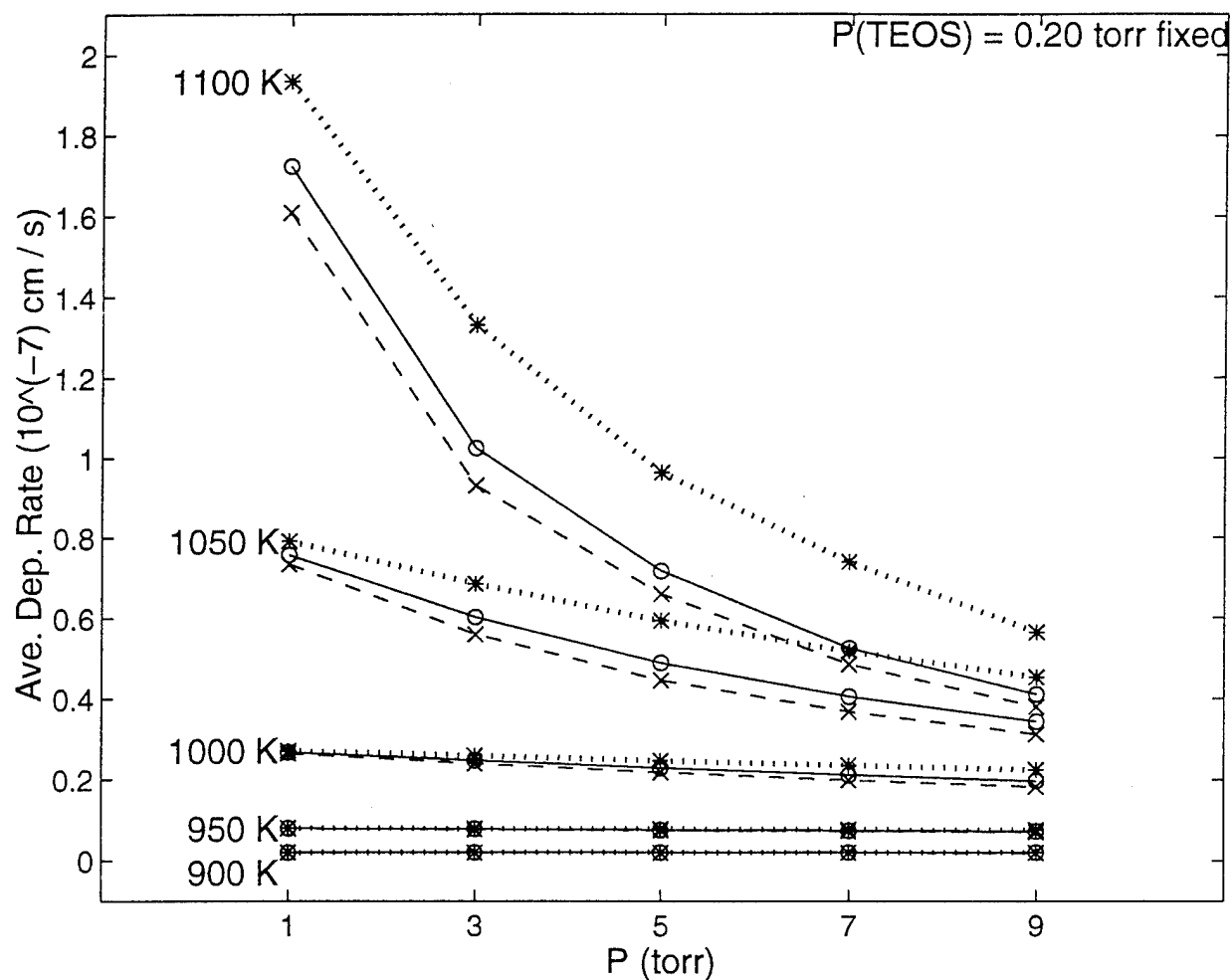
Figure 7: Average deposition rate per total area versus total pressure with temperature parameterized. Partial pressure of TEOS $P_{\text{TEOS}} = 0.20$ torr fixed.
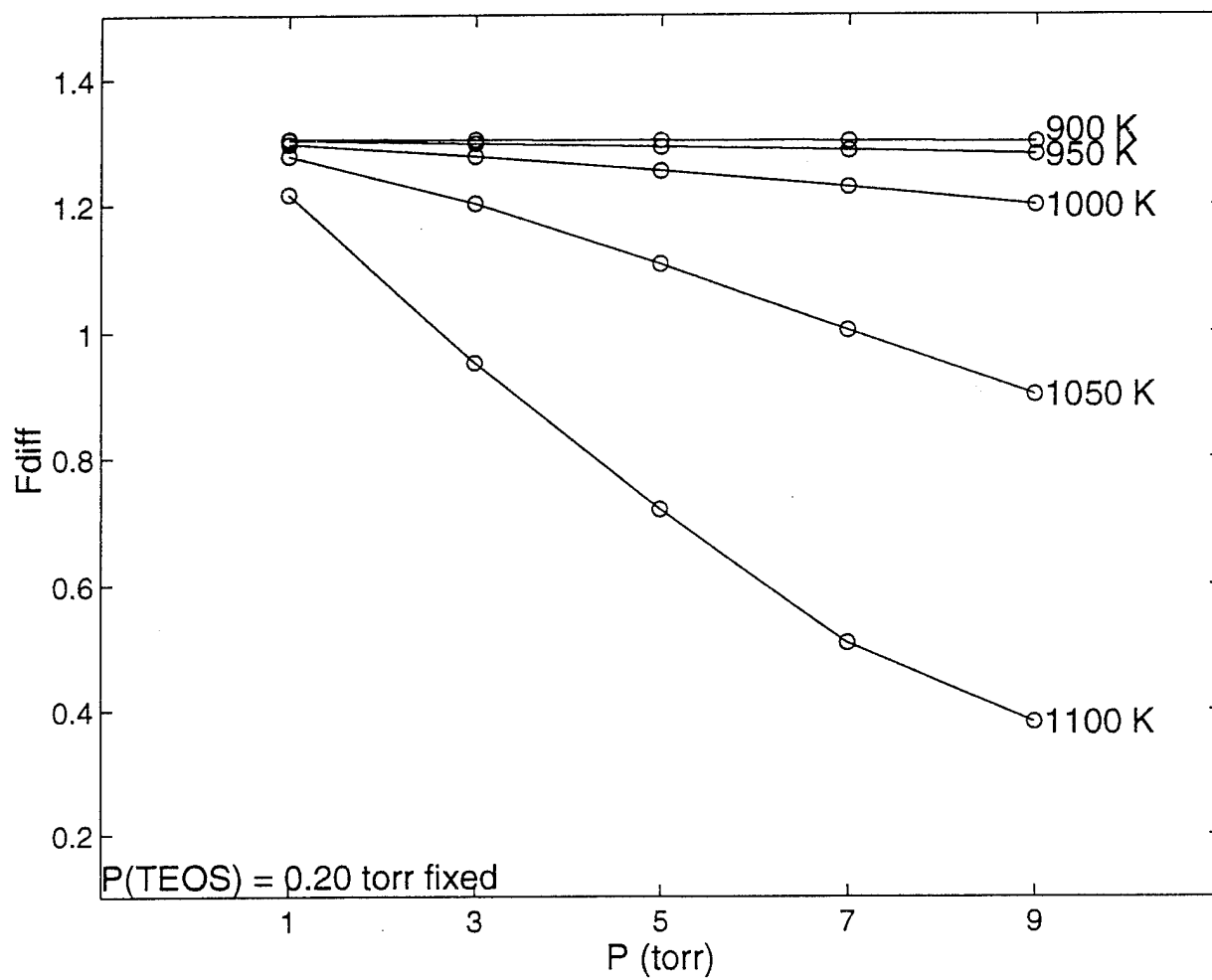
Figure 8: Fractional difference in the flux versus total pressure with temperature parameterized. Partial pressure of TEOS $P_{\text{TEOS}} = 0.20$ torr fixed.
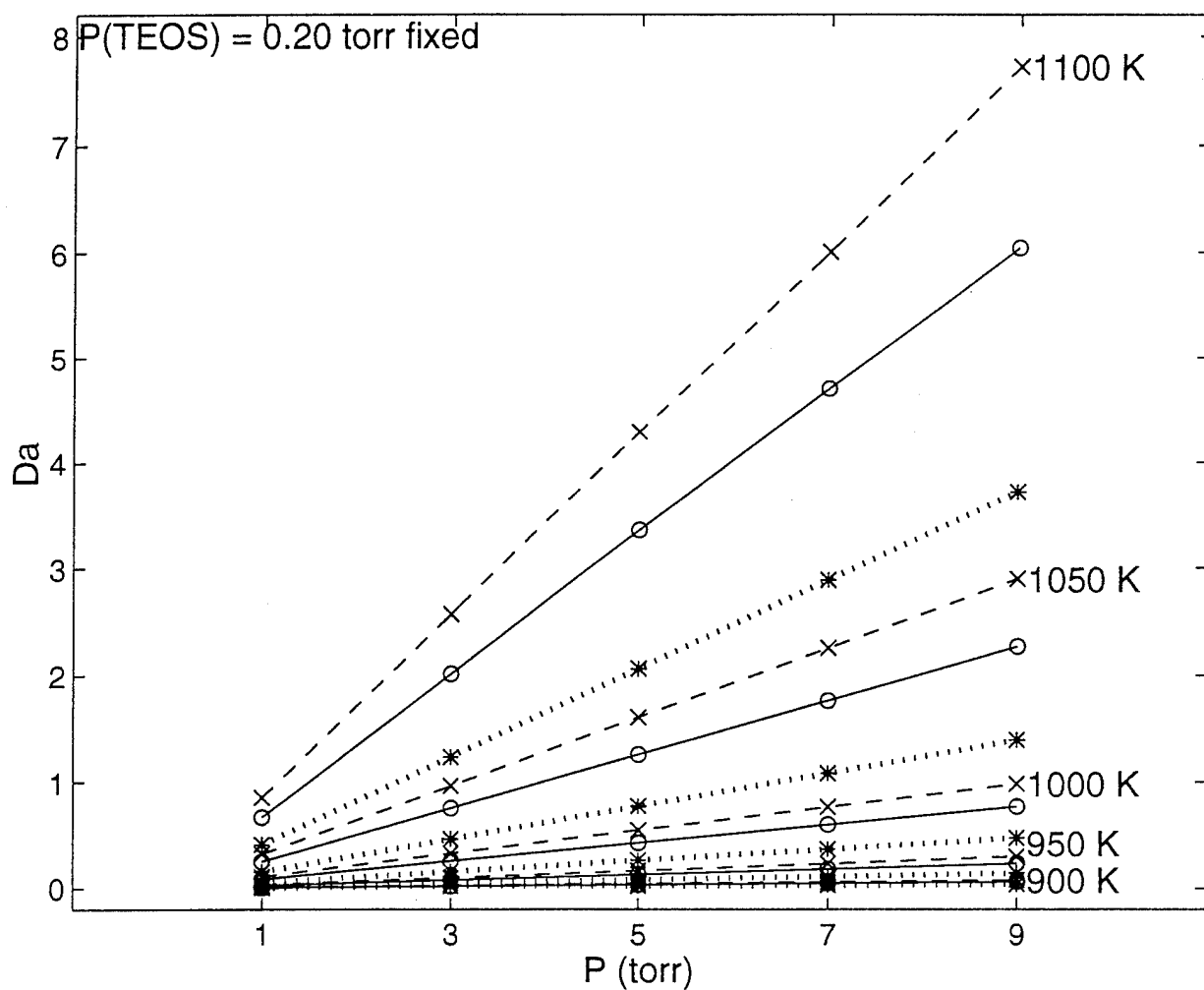
Figure 9: Generalized Damkoehler number versus total pressure with temperature parameterized. Partial pressure of TEOS $P_{\text{TEOS}} = 0.20$ torr fixed.

Figure 10: Effectiveness factor η versus Damkoehler number.

Figure 11: Arrhenius plot. Partial pressure of TEOS $P_{\mathrm{TEOS}} = 0.20$ torr fixed.

A HOMOGENIZATION TECHNIQUE FOR THE DEVELOPMENT OF MESOSCOPIC SCALE MODELS FOR CHEMICAL VAPOR DEPOSITION

Matthias K. Gobbert

Department of Mathematics, Arizona State University,
Tempe, AZ 85287-1804, gobbert@math.la.asu.edu

Doctoral Advisor: Dr. Christian A. Ringhofer
Defense Date and Thesis Approval Date: April 16, 1996
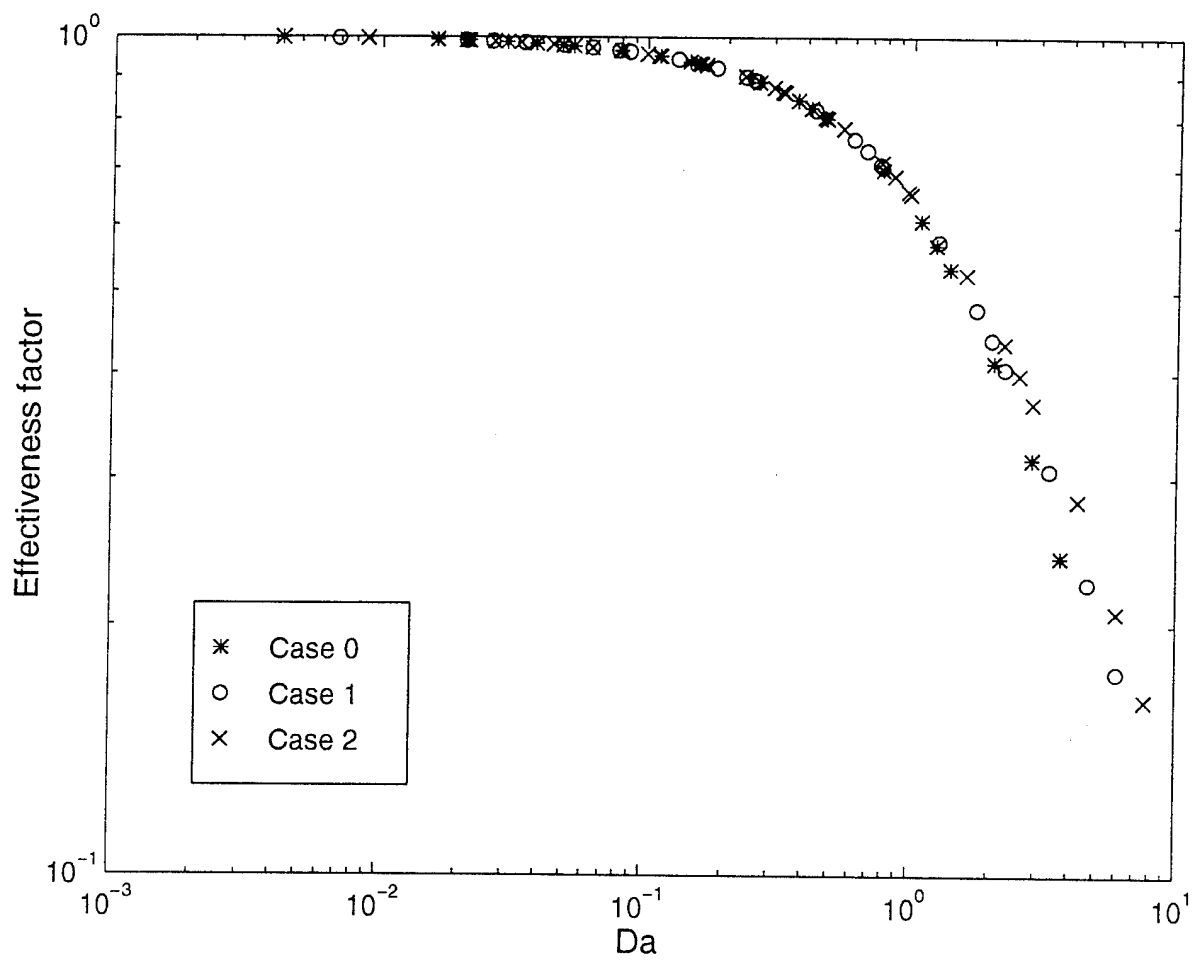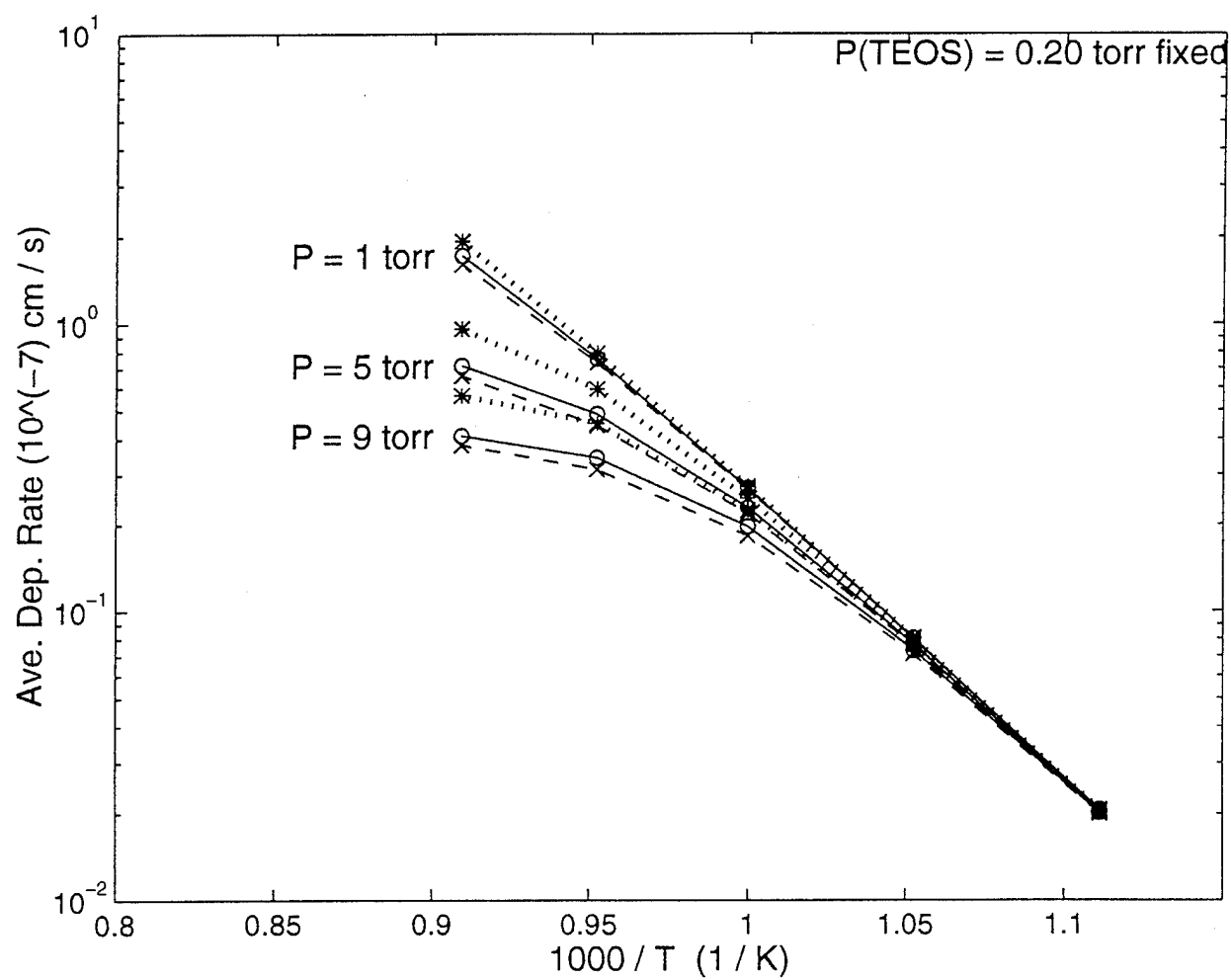AMS subject classifications: 65M06, 73B27, 76D30, 80A20, 80A32

Abstract

This dissertation presents a problem arising from the simulation of gas flow over microstructured surfaces. For the industrial application under consideration, the problem is appropriately given as a time-dependent nonlinear reaction-diffusion equation on a domain, which includes a flux condition on a boundary surface consisting of a microscopic fine structure. An equivalent problem for the bulk solution is derived, which incorporates all physical quantities of interest and which is accessible to efficient numerical simulation at the same time. This is achieved by applying a homogenization technique to the boundary condition, which eliminates the microscopic scale while retaining its effect on the bulk solution. The derivation presented in this dissertation is valid for a three-dimensional domain and a general boundary surface given in parameterized form.

The underlying application area in semiconductor manufacturing is the modeling of chemical vapor deposition in single wafer reactors. To the end of a global process model, this work introduces a mesoscopic scale model intermediate in length scale to the established reactor scale and feature scale models. The numerical simulation of this model is made possible by the homogenization technique above. The model is validated for a mathematical test problem by comparison to a classical numerical solution with full resolution of the fine structure. Furthermore, two studies on the physical example of thermally induced deposition of silicon dioxide from tetraethoxysilane (TEOS) are presented. The first study analyzes the effect of varying feature density, while the second one studies the influence of varying operating conditions on important physical parameters. A generalized Damkoehler number, which characterizes the amount of transport limitations in a way appropriate for the mesoscopic scale model, is introduced.

1

PROGRAMMED RATE CHEMICAL VAPOR DEPOSITION OF BLANKET TUNGSTEN THIN FILMS

Kathryn M. Tracy, Srikanth Bolnedi and Timothy S. Cale

Department of Chemical, Bio & Materials Engineering and Center for Solid State
Electronics Research, Arizona State University, Tempe AZ 85287

Executive Summary

Conventional, constant rate chemical vapor deposition (CRCVD) processes maintain constant pressure and temperature during processing. Programmed Rate CVD (PRCVD) has been proposed as a means of increasing wafer throughput relative to CRCVD processes, while maintaining step coverage and other critical film properties at acceptable values [1,2]. For PRCVD, the deposition rate is decreased as feature aspect ratio increases during feature fill in order to maintain a specified step coverage. The deposition rate is higher than the CRCVD rate at the beginning of film deposition, and is reduced as the aspect ratio increases towards the end of feature fill. The total time required to fill features with good step coverage is less than the time required by CRCVD processes because the average deposition rate is higher. We present early results in our experimental efforts to demonstrate PRCVD of tungsten.

Extended Abstract

The PRCVD process we use for blanket tungsten thin film deposition follows the protocol described by Cale *et al* .[1,2]; i.e., we decrease the temperature during deposition in order to decrease the deposition rate. Unpatterned silicon (100) wafers (125mm diameter, p-type) were used to study film characteristics such as thickness, density, grain size distribution, composition, and stress. All wafers were cleaned in 1% HF for 3 minutes prior to deposition. Blanket tungsten films were deposited in a Spectrum 202, cold wall, single wafer CVD reactor with a radiant heat source, using the hydrogen reduction of tungsten hexafluoride. Table 1 lists the conditions used for the first screening experiment, labeled (I), and Figure 1 shows the temperature trajectory .

Film thickness was determined for all samples by Scanning Electron Microscope (SEM). Film thicknesses were about 5 microns (Figure 2). Sheet resistance was measured using a four-point probe, and film resistivites were determined. The resistivities of the films (Figure 3) were relatively high compared to CRCVD samples, but within published values. Film adhesion to the silicon substrates was good.

Rutherford Backscattering Spectroscopy (RBS) was used to determine the film density and composition. The film thickness was first determined by the SEM, then RBS was performed. The density of sample I.3 was $16.1 g/cm^3$, compared to the bulk value of $19.2 g/cm^3$. The RBS spectrum for this sample (Figure 4) showed a carbon peak at the W-Si interface, probably due to contamination during the wafer cleaning procedure prior to film deposition. We were unable to detect oxygen or fluorine in this relatively thick tungsten film sample due to sensitivity impairment.

Auger Electron Spectroscopy (AES) was used to study the film composition of sample I.4, on the surface and as a function of depth. The AES spectra for the surface probe (not shown) indicated that the surface of the film had oxidized, as expected. Carbon was detected on the surface, possibly due to sample handling prior to the technique. The

AES depth profile (Figure 5) was limited to 100 nm for practical reasons, as the sputter rate was slow compared to the sample film thickness. AES did not show a significant oxygen content at the depth tested.

The microstructure of sample I.2 was studied using Transmission Electron Microscopy. A TEM cross section indicated that the grain size varied from 50 nm for the nucleation layer at the W-Si interface, to 0.5-1.0 micron at intermediate distances to the film surface. A non-uniform layer of tungsten silicide was found at the W-Si interface. Voids between the columnar tungsten grains were observed.

Residual stress measurements were performed on samples from a second experiment, labeled (II). Stress results were compared for films grown using the original PRCVD temperature trajectory (starting temperature 650°C), PRCVD with a reduced starting temperature (450°C), and CRCVD with the same temperature as the end-of-deposition temperature for both PRCVD samples (360°C). In all cases, the total pressure was 1.6 Torr, H_2 flow rate was 500 sccm, and WF_6 flow rate was 50 sccm. The high temperatue PRCVD film (II.2) had significantly lower stress (Table 2) than the other two samples.

The results from the first set of experiments (I.*) provide information about the composition and structure of PRCVD tungsten films on silicon substrates. The PRCVD samples I.3, I.4 and I.5 were compared to CRCVD films deposited at the same pressures, flow ratios and flow rates. The densities of the PRCVD films obtained to date are lower, apparently due to voids formed at the Si-W interface. These voids may not form in the presence of a conventional nucleation layer, such as TiN. The resistivities of our films were relatively high, which may be due to smaller average grain sizes, and the lower film densities. The motivation for continuing to test this protocol is apparent when the time of deposition is compared. According to this study, the PRCVD process will improve throughput by at least 40% compared to CRCVD, with equivalent step coverage predicted [1,2]. If film properties and step coverage are good when a nucleation layer is used, the PRCVD process would provide improved throughput and decreased consumables.

1. T.S. Cale, M.K. Jain and G.B. Raupp, in Tungsten and Other Advanced Metals for ULSI/VLSI Applications V, S.S. Wong and S. Furukawa, eds., Materials Research Society, 1990, p. 179.
2. T.S. Cale, M.K. Jain and G.B. Raupp, J. Electrochem. Soc., 137(5), 1526 (1990).

Figure 1: Experimental Conditions for PRCVD screening runs; P = 0.5 Torr. See Figure 1 for Temperature Trajectory.

Figure 2: Stress Measurements for CRCVD and PRCVD. P = 1.6 Torr, H_2 = 500 sccm, WF_6 = 50 sccm.

ID #	H_2 :WF_6	Total Flow
I.1	10	176 sccm
I.2	10	352 sccm
I.3	5	180 sccm
I.4	10	528 sccm
I.5	10	704 sccm

Run	Temperature Profile	Stress (Dyn/cm^2)
II.1	CRCVD, 360°C	1.2 E10
II.2	PRCVD, 650-360°C	3.0 E9
II.3	PRCVD, 450-360°C	1.1 E10

PROGRAMMED RATE CHEMICAL VAPOR DEPOSITION:

BLANKET TUNGSTEN FILM CHARACTERIZATION

by

Kathryn M. Tracy

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

ARIZONA STATE UNIVERSITY

May 1996

ABSTRACT

Improving wafer throughput for single-wafer, low pressure chemical vapor deposition (LPCVD) reactors is problematic, since process changes needed to achieve high throughput often lead to film property degradation. Conventional LPCVD process in general do not have enough degrees of freedom to achieve high throughput and maintain good film properties. In this work, the film properties resulting from a new process protocol, programmed rate chemical vapor deposition (PRCVD), are investigated. Blanket tungsten (W) film deposition, based on the hydrogen reduction of tungsten hexafluoride (WF_6), is used as a test vehicle for this new protocol. Results are compared to conventional, constant rate CVD (CRCVD) thin films produced with the same equipment.

Results from this study of blanket tungsten PRCVD show consistent, excellent adhesion to silicon substrates and deposited on nucleation layers using the silane reduction of WF_6. The protocol is shown to be viable and results in significant time savings. Step coverage is shown to be equivalent to CRCVD samples using the same equipment. High density tungsten thin films with low resistivity have been produced using the PRCVD process, and the films produced show significantly lower stress than the CRCVD samples. Finally, SIMS results showed that the one PRCVD film analyzed had significantly lower fluorine concentration compared to the CRCVD control, indicating that PRCVD process may exhibit improved chemical purity.

PROGRAMMED RATE CHEMICAL VAPOR DEPOSITION OF TUNGSTEN

John Kristof, Kathy Tracy, LiJuan Song, Kostas Tsakalis, Timothy Cale
Chemical Bio & Materials Engineering,
Arizona State University, Tempe AZ 85287-6206
and
Gerrit Leusink
Materials Research Corporation
Gilbert, AZ 85233

ABSTRACT

For deposition processes, the goal is to maximize throughput while maintaining acceptable film properties. Higher deposition rates in general lead to poorer step coverages due to reactant depletion caused by the higher reaction rates, or due to depletion along the depth of the features. Using feature scale simulations, we have proposed a protocol called programmed rate CVD (PRCVD). PRCVD overcomes increasing aspect ratios as features approach closure during deposition. The initial rate can be much higher than conventional rate CVD (CRCVD) and is decreased during the deposition.

We tested the PRCVD concept for blanket tungsten deposition from tungsten hexafluoride and hydrogen, in a lamp-heated, cold-wall, single wafer, LPCVD reactor. We found significant time savings (PRCVD took 70% less time than CRCVD processes which give the same step coverage), resistivities and stresses were equivalent to or better than CRCVD processes. Thus, it can be concluded that PRCVD is a viable process for tungsten deposition.

PRCVD PROCESS vs CRCVD PROCESS

CRCVD maintains constant temperature, flow rates and pressure. In order to improve throughput for CRCVD, the deposition rate can be increased; however, this can lead to poor step coverage. This limitation is due to using fixed deposition conditions, although feature aspect ratios increase during deposition.

PRCVD allows for programming deposition conditions to add degrees of freedom. With PRCVD, deposition rate is decreased during deposition as feature aspect ratios increase. This allows the initial rate to be much higher than the average rate, perhaps as much as an order of magnitude higher. A major concern in using the PRCVD process would be the effect on film properties: uniformity, stress, resistivity, adhesion, composition. [1,2]

Experimental Design

The reactor we used is a Spectrum 202 Single Wafer Reactor. The deposition chemistry is hydrogen reduction of tungsten hexafluoride,

$$WF_6 + 3H_2 \rightarrow W + 6HF$$
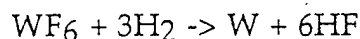
according to the hetrogeneous reaction kinetics,

$$R = k_o \cdot exp(\frac{-E}{RT})(\frac{P_{H_2}^{1/2} P_{WF_6}}{1 + K_{WF_6} P_{WF_6}})$$

Experiments were carried out on unpatterned, bare silicon wafers, unpatterned and patterned 50 nanometer sputtered titanium nitride silicon dioxide wafers. Wafers were passed through an isopropyl alcohol clean after an HF (50:1) dip prior to processing. After processing, they were measured for weight gain, resistivity, uniformity, stress and composition.

PRCVD Control Construction

Compared to CRCVD, PRCVD controls one or more critical process parameters. In this work, we controlled the temperature ramp rate, as seen in a typical program in Figure 1.
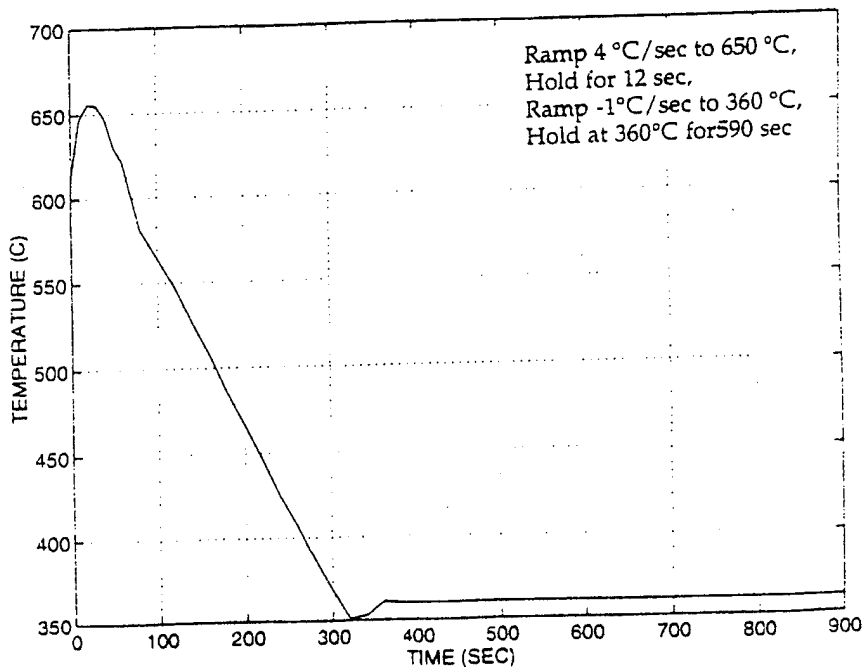
Figure 1. Temperature Ramp, PRCVD

The temperature ramp was achieved using a closed loop Eurotherm Controller. Compared to a constant temperature CRCVD, where the process starts and ends at the same temperature, the PRCVD process begins at an initial higher temperature and is ramped downward as feature closure increases aspect ratio.

Using this control, we achieved time savings from 50 - 70%, using the following formula, [3]

2

$$Time\_Saved\_\% \equiv [\frac{t_{crcvd} - t_{prcvd}}{t_{crcvd}}] \times 100\%$$

The SEM results shown here are for 450 $^{\circ}$C, which is compatible with the lower temperatures required for aluminum. Our work has tested initial starting temperatures of 450°, 550°, and 650° C, with temperature ramps ranging from -1 to -5 $^{\circ}$C / second.
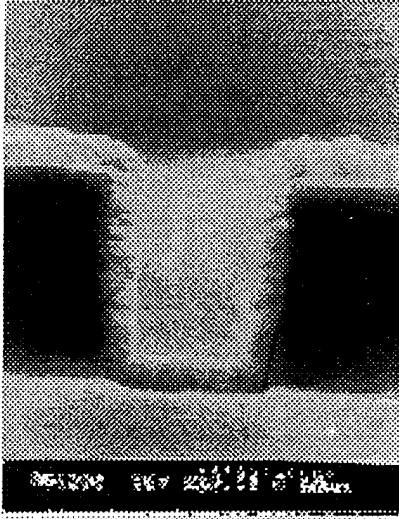
Figure 3. SEM of PRCVD results, t = 90 s, 0.5 T, T= 450 $^{\circ}$C
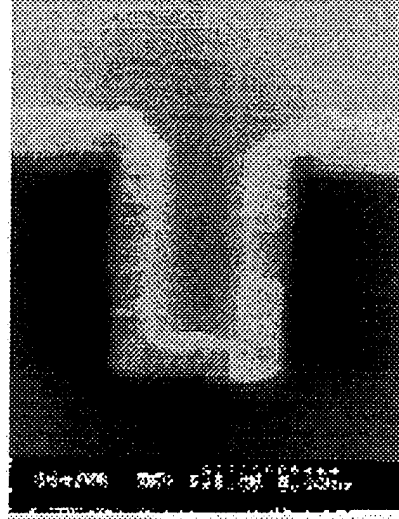
Figure 4, SEM of CRCVD results, t=465 s, P=0.5, T = 360 $^{\circ}$C

Run	Process (Pressure, Torr, Temp,$^{\circ}$C)	Aspect Ratio	Thickness at surface (t,micron)	Thickness at side-wall (B, micron)	Step Coverage (B / t)	Deposition Time (seconds)
1	PRCVD (2, 450)	1.60	0.12	0.12	1.0	90
2	CRCVD (2, 360)	1.58	0.10	0.10	1.0	465
3	PRCVD (0.5, 450)	1.20	0.16	0.15	0.94	90
4	CRCVD (0.5, 360)	1.25	0.19	0.19	1.0	465
5	PRCVD (5, 360)	1.67	0.29	0.27	0.93	90

3

6	CRCVD (5, 360)	1.36	0.32	0.28	0.88	465
7	PRCVD (2, 450)	1.45	0.26	0.21	0.81	90

Table 1: Experiments on Patterned TiN Wafers

Experimental Results

Secondary Ion Mass Spectroscopy (SIMS) analysis revealed that PRCVD shows promise of lower fluorine incorporation than CRCVD. Residual stress increased with increasing pressure, as did weight gain and hence density. Higher pressure also decreased non-uniformity.

Uniformities are about the same for PRCVD and CRCVD; we achieved step coverages approaching and equal to SC = 1 (where SC is the ratio of sidewall coverage to surface coverage) with considerable time savings by using PRCVD.

Conclusions

PRCVD protocol simulations using EVOLVE (a low pressure deposition simulator) theoretically predicted tungsten step coverage equivalent to CRCVD processes. A major concern for using PRCVD would be to sustain film properties at CRCVD levels, with improvement or at the very least without degradation. Here, we have shown an 82% time savings, while preserving critical wafer parameters of stress, uniformity, resistance and density. Thus, it can be concluded PRCVD is a viable process for tungsten deposition.

References

T.S. Cale, M.K. Jain and G. B. Raupp, in Tungsten and Other Advanced Metals for ULSI/VLSI Applications V, S.S. Wong and S. Furukawa, eds., Materials Researcch Society, 1990, p. 179.

T.S. Cale, M.K. Jain and G.B. Raupp, J. Electrochemical Society, 137(5), 1526 (1990).

K.M. Tracy, "Programmed Rate Chemical Vapor Deposition: Blanket Tungsten Film Characterization", Master's Thesis, Arizona State University, 1996.

APPLICATION OF FEED-FORWARD AND ADAPTIVE FEEDBACK CONTROL TO SEMICONDUCTOR DEVICE MANUFACTURING

K. Stoddard, P. Crouch, M. Kozicki and K. Tsakalis

Arizona State University

Center for Systems Science and Engineering

Box 877606, Tempe, AZ 85287-7606

Abstract

A feed-forward and adaptive feedback control methodology is developed and experimentally applied to several different processes commonly used in the fabrication of semiconductor integrated circuit devices. A circular parallel-plate capacitor with a glass (SiO_2) dielectric is manufactured on silicon wafers to illustrate the use of these control strategies in the processes of silicon oxidation, aluminum metallization, lithography, and aluminum etching. The goal is to maintain a constant capacitance value on a run to run basis regardless of disturbances or modeling errors in the processes.

1 Introduction

As we move into the twenty-first century, the semiconductor industry will require processing equipment with the capability to fabricate devices with 0.2 micron features. To manufacture devices of this size, it is necessary to maintain very tight processing tolerances that, in turn, will inevitably require the application of suitable control methodologies to these processes.

Recent studies have demonstrated the potential of feedback control laws in maintaining the desired process characteristics for several types of processes encountered in semiconductor manufacturing. These include real-time control strategies aiming to maintain a constant processing environment [1, 2, 3, 4, 5], run-to-run strategies aiming to adjust the process inputs so as to obtain the desired product specifications [6, 7, 8] and monitoring/diagnostic/control schemes [9].

In this study we adopt a run-to-run feed-forward and adaptive feedback control strategy and implement it for the overall sequence of processing steps in the experimental manufacturing of a parallel plate capacitor. This processing sequence involves four basic steps which are commonly used in semiconductor manufacturing processes, namely, silicon oxidation, aluminum metallization, lithography and aluminum etch. In the first step (silicon oxidation), the glass dielectric or oxide layer of the capacitor is grown on a bare silicon wafer. In the second step (aluminum met-

allization), the aluminum is deposited over the oxide grown in the first step which is the material of the top plate of the capacitor. In the third step (lithography), the photoresist pattern of the top capacitor plate is formed over the aluminum layer. In the fourth and final step (aluminum etch), the aluminium that is not protected by the photoresist is etched away to physically form the top plate of the capacitor. After completing each of these steps, a suitable output of the process is measured in order to assess the partial state of the wafer and determine the control actions to be taken. The objective is to maintain a constant, desired value of the capacitance which can be found by means of an electrical measurement after the completion of all four steps.

From static electric field theory, the capacitance value for a parallel-plate capacitor can be expressed as [10]

$$C = \frac{Q}{V_{12}} = \epsilon \frac{S}{d} \qquad (1)$$

where Q is the absolute charge on the capacitor plates, V_{12} is the potential difference between the two capacitor plates, ϵ is the permittivity constant of the dielectric, S is the area of the capacitor plates, and d is the dielectric thickness. The relationship between the parallel plate area and the dielectric thickness provides us with the opportunity to apply feed-forward control to maintain a constant capacitance value for a variable dielectric width by simply adjusting the capacitor plate area. In this study, feed-forward control is used in both the lithography and aluminum etch steps of the capacitor fabrication process.

On the other hand, the development of a simple, run-to-run, adaptive feedback control law has been studied for a silicon oxidation process to improve performance and compensate for equipment drifts and modelling errors [11]. In order to illustrate its versatility, we apply the same controller design principle, to a physically based model for the silicon oxidation step and an empirical model for the aluminum etch step. The parameters of the physically based model used in the silicon oxidation step are initially chosen far from the actual parameter values of the system to demonstrate the adaptive capability of the controller when the actual model parameters are not well
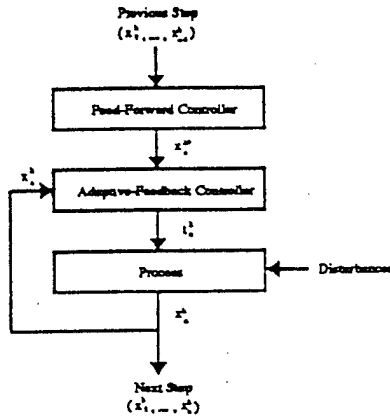
Figure 1: Typical Feedforward/Feedback Controller Configuration.

known. This selection also serves to demonstrate the corrective capability of the feed-forward controllers in the lithography and aluminum etching steps. In both cases, the adaptive-feedback controllers have the ability to compensate for, at least limited, process drifts while preserving the performance of the system.

This paper is organized as follows. The feed-forward and adaptive feedback controller design methodology is discussed in Section 2. Section 3 contains a description of the fabrication steps of the capacitor and some details regarding the control of these processes. The experimental results are presented in Section 4 while Section 5 contains concluding remarks and directions for future research.

2 Control Methodology

The feed-forward and adaptive feedback control methodology used in this study consists of two separate controllers that are used in conjunction with each other or separately depending upon the application. A typical configuration of the feed-forward and adaptive feedback controllers, relative to a given process step, is shown in Fig. 1 and further details on their design are discussed below.

2.1 Adaptive-Feedback Controller
Following [11], a simple, run-to-run, adaptive feedback controller can be used to compensate for certain types of modeling errors as well as process drifts. The implementation of this controller relies on a parametric model of the process. The derivation of such a model can be either physically based or empirical, i.e., based on data obtained experimentally from the system. The parameters obtained off-line from a physically based mathematical model of the process are at best an approximation, often requiring "fine-tuning" to improve the accuracy of the description. On the other hand, the parameters of an empirical model are typically obtained from a least-squares fit of several input-output data points to some suitable function. In both cases, the development of an accurate model requires extensive and expensive experimentation. This can be partially avoided by performing the

final fine-tuning of the model parameters by means of an on-line adaptive scheme.

For example, assuming that f is a known, locally well behaved function that describes the process (i.e., for some θ_*, the input-output pairs (u, x) around an operating point satisfy $u = f(x, \theta_*)$), the adaptive feedback controller is given as

$$u_n^k = f(x_n^{k*}, \theta_n^k) \qquad (2)$$

where u_n^k is the control, θ_n^k is the vector of the parameter estimates and x_n^{k*} is the desired output of the process. The subscript n denotes the station number and the superscript k denotes the wafer or run number. The parameter estimates are then updated based on the actual measurements so as to achieve convergence of the output to (or near) the desired set-point as $k \to \infty$. From [11] such an adaptive law is

$$\theta_n^{k+1} = \theta_n^k + \gamma_n^k(x_n^{k*} - x_n^k) \qquad (3)$$

where θ_n^{k+1} is the updated model parameter that will be used for the next wafer, x_n^k is the output measurement (feedback) of the process, and γ_n^k is defined as

$$\gamma_n^k = \frac{\partial f^\dagger}{\partial \theta}(x_n^{k*}, \theta_n^k) \left(\frac{\partial f}{\partial x}(x_n^k, \theta_n^k) \cdots \right.$$
$$\left. + \frac{\partial^2 f}{\partial \theta \partial x}(x_n^k, \theta_n^k)\frac{\partial f^\dagger}{\partial \theta}(x_n^k, \theta_n^k)[u_n^k - f(x_n^k, \theta_n^k)] \right)$$

where $x^\dagger \triangleq x^T[xx^T]^{-1}$ for a row vector x. Notice that to ensure that such an adaptive scheme is "well-behaved" in the presence of inevitable modeling errors and measurement noise, suitable dead-zone and parameter projection modifications should be incorporated in the adaptation [12]. These issues are not discussed here since, due to the limited number of experiments, they are of minor consequence.

2.2 Feed-Forward Controller
The feed-forward controller is used to determine the desired output of the current step from the results obtained from the previous processing steps. Its objective is to adjust the desired output of the current step so as to partially compensate for errors occurring in any of the previous steps. This controller is typically a simple expression which provides a relationship between the output of upstream processing steps and the desired output of the current step or downstream processing steps, i.e.,

$$x_n^{k*} = h(x_1^k, \ldots, x_{n-1}^k) \qquad (4)$$

where h is a known function, x_n^{k*} is the desired output of the current processing step and $x_1^k, \ldots, x_{n-1}^k$ are the output measurements of the previous processes. Although in our study the design of the feed-forward controllers is rather straightforward (see the following section for details), notice that in general, the computation of h may be quite involved and require the on-line solution of an optimization problem.

3 Fabrication Process

A four step fabrication process was employed to manufacture a circular parallel-plate capacitor with a SiO$_2$ dielectric. Each of these fabrication steps is discussed in the following subsections along with the implementation of the feed-forward and adaptive-feedback control methodology.

3.1 Silicon Oxidation Step

The first step in the fabrication process is the growth of an oxide layer on the wafer surface using a Rapid Thermal Processing (RTP) system. This oxide layer is the dielectric of the capacitor whose thickness directly determines the capacitance value of the device making it necessary to maintain a consistent thickness. To achieve this, an adaptive feedback controller was applied to the RTP system.

This adaptive feedback controller uses the oxidation time as the manipulated variable to control the oxide thickness grown on the wafer. Other variables such as temperature, oxidant type, oxidant flow rate, and wafer crystal orientation are all held constant, e.g., by means of local PID loops. Note that there was no special tuning performed on the local PID loops. This is due to the flexibility offered by the adaptive feedback controller in the outer loop, requiring only the consistent (precise) but not necessarily accurate operation of the inner loops. Further, an RCA clean of each wafer was performed before growing the oxide layer to provide a somewhat constant native oxide thickness to reduce noise that could adversely effect the controller.

The Deal-Grove oxidation model was used to describe the oxidation process occurring in the RTP system [13]. The resulting control law equation is a second order polynomial containing three parameters. The initial model parameters for this physically based model were chosen far from the actual parameter values of the RTP system for two purposes. First, to create a large disturbance in the first step of the fabrication process to illustrate the capability of the feed-forward control algorithms in the lithography and etching steps. Second, to demonstrate the adaptive capability of the controller when the initial guess of the parameters is poor.

3.2 Aluminum Metallization Step

The second step in the fabrication process is the evaporation of aluminum over the oxide layer using a electron beam metal evaporation system. This aluminum layer makes up the top plate of the capacitor. The thickness of the aluminum does not directly affect the capacitance value of the device; it does, however, indirectly affect the amount of time required to completely etch through the aluminum to form the top plate of the capacitor in the aluminum etch step. Since such thickness variations can effectively act as a a source of noise in the control of the aluminum etching step, it is important to maintain a constant aluminum thickness at this stage. The electron beam metal evaporation system is equipped with a computer which monitors the metal deposition rate and total deposition on the wafer. With this control capability, it is not necessary to use any other control scheme to maintain a consistent aluminum thickness

in this fabrication step.

3.3 Lithography Step

The third step in the fabrication process is the creation of the resist pattern of the top plate of the capacitor over the aluminum layer. This step consists of the following five sub-steps.

The first sub-step is the application of positive photoresist over the aluminum layer using a spin coating system. The second is the soft-bake or pre-bake of the wafer which involves baking the wafer for 30 minutes at 80°C to remove the solvent from the photoresist so that it is properly exposed in the following step. The third sub-step is the exposure of the mask image of the top capacitor plate in the photoresist using a stepper projection printer. The fourth sub-step is the development of the photoresist which removes the photoresist that was exposed in the third sub-step to create the resist pattern on the top plate of the capacitor. The fifth and final sub-step is the hard-bake or post-bake of the wafer which amounts to baking the wafer for 10 minutes at 110°C to provide better adherence of the resist to the wafer surface and, also, an increased resistance to the etch process.

At this point, based on practical experience, it was decided that a run-to-run feedback controller for the lithography steps would be unnecessary. Instead, each of these sub-steps was experimentally optimized so as to maintain consistent processing results between runs. The exposure sub-step, however, does provide us with the first real opportunity to apply feed-forward control to this fabrication process. The stepper printer can be intentionally defocused to increase or decrease the size of the capacitor plate image on the mask. With this capability, it is possible to compensate for small errors in the oxide thickness grown in the first step by increasing or decreasing the size of the top plate of the capacitor. The feed-forward controller was derived from equation (1) to determine the amount of defocus required to attempt compensation for any error in the measured oxide thickness.

3.4 Aluminum Etching Step

The fourth step in the fabrication process is the etching of the aluminum layer to form the top plate of the capacitor using a Reactive Ion Etching (RIE) system. The aluminum etching chemistry uses Cl$_2$ and BCl$_3$ which etches at a 5:1 ratio between the aluminum and photoresist. The RIE process has the desirable property that it etches both vertically and horizontally. With the ability to etch horizontally, it is possible to adjust the area of the top plate of the capacitor to compensate for errors in the oxide thickness as well as in the size of the resist pattern of the capacitor plate using controlled overetching. This correction capability, however, is limited in that it is only possible to reduce but not increase the size of the top plate of the capacitor (it is not physically possible to un-etch the aluminum plate).

In this step, both feed-forward and adaptive feedback control were used. The feed-forward controller was again derived from equation (1) to determine the amount of radial overetching required to obtain a plate area that maintains a constant capacitance value for the oxide thickness grown in the first fab-
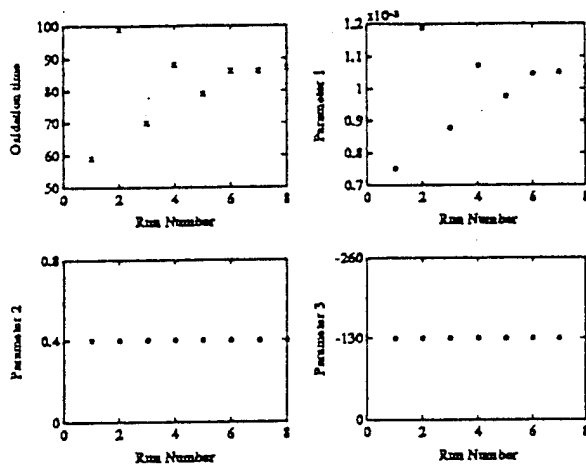
Figure 2: Oxidation time and parameter estimates.

Figure 3: Desired vs. achieved oxide thicknesses.

Figure 4: Desired, corrected and mask plate areas in the lithography step.

rication step. An empirical feature scale model was developed to predict the radial overetching capability of the RIE process with respect to time. All other variables such as RF power, gas flow rates, and pressure were held constant. The resulting radial overetch model is parabolic, requiring a second order polynomial with three parameters to describe [13]. These three parameters were determined using a least-square fit on the experimental data obtained off-line from the RIE system.

4   Experimental Results

In this section, we discuss the performance of the feed-forward and adaptive-feedback controllers, applied to the silicon oxidation, lithography, and aluminum etching steps.

In the silicon oxidation step, we applied an adaptive feedback controller to the RTP system. The model parameters were chosen in such a way as to illustrate the adaptive capability of the controller as well as to introduce a disturbance into the fabrication process to show the feed-forward capability in the lithography and aluminum etching steps. The response of the oxidation time and the three parameter estimates generated by the controller are shown in Figure 2. The model parameters and oxidation time converge rapidly within six samples. The resulting oxide thicknesses for each wafer (x) and the corresponding desired value (o) are shown in Figure 3. After the initial adaptation transient, the oxide thickness remains within 0.5 percent from the desired value. The main causes of this variation are that the oxidation time is rounded to the nearest whole second and the measurement of the oxide thickness is accurate to within two angstroms.

In the lithography step, we applied a feed-forward controller to adjust the size of the resist pattern on the top plate of the capacitor. This controller aims to compensate for small variations in the oxide thickness by adjusting the focus of the stepper printer. Because of apparatus-related limitations, the values
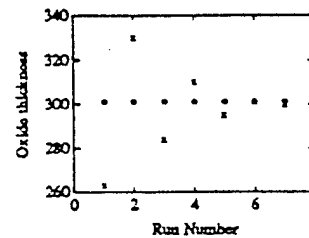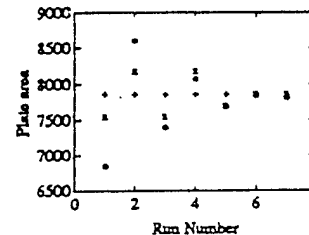
of such an adjustment are limited to only ±4% or ±2% of the resist pattern area. Therefore, it is not possible to completely correct large oxide thickness errors or correct the plate area to the exact size that is required. After this step, it is necessary to overetch in order to obtain the exact plate area desired. To assess the performance of the controller in this step, the corrected or defocused plate area (x) the desired plate area (o) and the actual plate area of the image on the mask (+) are shown in Figure 4. Note that due to the limitations of the apparatus, there is little to be done about the large oxide thickness errors occurring in the first three wafer runs. In the later runs, however, the smaller oxide thicness errors are easier to correct.

In the aluminum etching step, we used both feed-forward and adaptive feedback controllers to reduce the area of the top plate of the capacitor by overetching the aluminum layer. The response of the etch time is shown in Figure 5.a. (The model parameters remain virtually unchanged during adaptation, since their initial guess —obtained by an off-line least-squares fit— was fairly accurate.) The fluctuations of the etch time show the amount of radial overetching required to adjust the plate area. In the second pro-
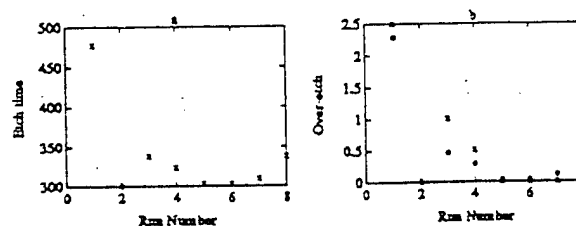
Figure 5: Aluminum Etch step: a. Etch time; b. Actual and desired overetch.

Figure 6: Capacitance values: a. Calculated vs. desired; b. Measured.

cessing run, the plate area needed to be larger than the resist pattern, so the controller performed a minimal etch to form the capacitor plate. This is due to the fact that the plate area can only be reduced through overetching the aluminum layer. The resulting radial overetch measurement for each wafer (x) and the desired radial overetch value (o) are shown in Figure 5.b. It should be mentioned at this point that there is an inherent performance limitation in this step since the measurement of the radius of the capacitor plate is only accurate to within ± 0.5 microns.

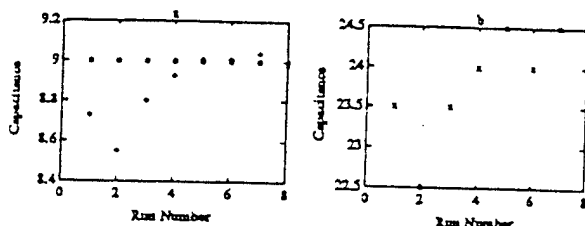Finally, the capacitance value was calculated from the resulting measurements of the oxidation layer and the top plate area of the capacitor. A comparison between the calculated capacitance (+) and the desired capacitance value (o) for each wafer run is shown in Figure 6.a. After the initial adaptation transient, the resulting capacitance values had a maximum variation of 0.5 percent from the desired value. However, when the actual capacitance of these devices was measured, the results were significantly different from the calculated capacitance values. This deviation is attributed to the electric field fringing effects that are neglected in the derivation of equation (1). In the fabricated device, the contribution of fringing is rather significant by increasing the effective area of the top capacitor plate and hence the capacitance. One possible remedy of this problem is to "close" the outer loop by using an adaptive feedback controller to adjust to the dielectric permittivity constant based on the values of measured and calculated capacitance. This approach was not implemented in this study because the electrical measurements were performed after the completion of all processing steps for all the wafers. Note that the validity of this argument becomes evident since the ratio of measured versus calculated capacitance is fairly constant: its average value is 2.68 with standard deviation 0.028. Nevertheless, it should be emphasized that the overall control scheme achieved a fairly consistent capacitance value (within 1 percent) despite the limitations of the various measurment devices.

5 Conclusions

The present study demonstrates the potential of using feed-forward and adaptive feedback run-to-run control techniques to manufacture semiconductor devices with desired characteristics. In the experi-

mental application of this approach, run-to-run controllers were implemented in almost every step of the complete manufacturing process with very encouraging results. In order to establish the viability of such a control scheme, further studies should adrress the issues of the controller behavior under more adverse conditions (e.g., "accidental" de-tuning of a process step, changes in operating conditions, processing drifts etc.) and for larger numbers of processed wafers.

Acknowledgments

References

[1] T. Breedijk, T. Edgar and I. Tvachtenberg, "A Model Predictive Controller for Multivariable Temperature Control in Rapid Thermal Processing," *Proc. 1993 ACC*, San Francisco, CA.

[2] S.A. Norman and S.P. Boyd, " Multivariable Feedback Control of Semiconductor Wafer Temperature, *Proc. 1992 ACC*, Chicago, IL.

[3] C.D. Schaper, "Real-Time Control of Rapid Thermal Processing Semiconductor Manufacturing Equipment," *Proc. 1993 ACC*, San Francisco, CA.

[4] M. Elta *et al.*, "Applications of Control to Semiconductor Manufacturing: Reactive Ion Etching," *Proc. 1993 ACC*, San Francisco, CA.

[5] A. Srinivasan, C. Batur and R. Veillette, "Projective Control Design for Multi-Zone Crystal Growth Furnace," *Proc. 1993 ACC*, San Francisco, CA.

[6] R.A. Soper, D.A. Mellichamp and D.E. Seborg, "An Adaptive Nonlinear Control Strategy for Photolithography," *Proc. 1993 ACC*, San Francisco, CA.

[7] S. Watts Butler and J. Stefani, "Application of Predictor Corrector Control to Polysilicon Gate Etching," *Proc. 1993 ACC*, San Francisco, CA.

[8] E. Sachs, A. Hu and A. Ingolfsson, "Run by Run Process Control: Combining SPC and Feedback Control," submitted to *IEEE Trans. Semicond. Manuf.*, Oct. 1991.

[9] C.J. Spanos, S. Leang and S. Lee, "A Control and Diagnosis Scheme for Semiconductor Manufacturing," *Proc. 1993 ACC*, San Francisco, CA.

[10] D. Cheng, *Field and Wave Electromagnetics*, Addison-Wesley, New York, 1989.

[11] K. Tsakalis and P. Crouch, "A Simple Adaptive Controller for an Oxidation Process," *Proc. 1992 ACC*, Chicago, IL.

[12] G.C. Goodwin and K.S. Sin *Adaptive Filtering Prediction and Control*, Prentice Hall, Englewood Cliffs, NJ, 1984.

[13] W. Runyan and K. Bean, *Semiconductor Integrated Circuit Processing Technology*, Addison-Wesley, New York, 1990.

Set-Membership Estimation for Weakly Nonlinear Models: An Application to the Adaptive Control of Semiconductor Manufacturing Processes*

Kostas S. Tsakalis and Lijuan Song

Arizona State University
Center for Systems Science and Engineering
Box 877606, Tempe, AZ 85287-7606
E-mail: tsakalis@enuxsa.eas.asu.edu

Abstract

In this paper we consider an application of set-membership concepts to the parameter estimation problem for weakly nonlinear models. We develop a recursive algorithm that, given input-output data, a bound on the measurement noise and a local bound on the Hessian of the nonlinear model with respect to the unknown parameters, produces a consistent ellipsoid containing the "actual" model parameters. To illustrate the use of this algorithm, we consider the process of oxidation of silicon in dry oxygen where the oxidation time is determined by means of a simple adaptive controller. In an effort to reduce the parametric uncertainty, we employ an auxiliary set-membership estimator to update the set of parameter constraints on-line and, thus, avoid unnecessary drifts of the adaptive controller parameters.

1 Introduction

Gate oxidation is one of the frequently encountered and critical process steps in any mainstream metal oxide semiconductor integrated circuit fabrication. Briefly described, the gate oxidation step is performed as follows: A batch of silicon wafers are introduced in an oxidation chamber and are oxidized at high temperatures in an oxygen atmosphere for a predetermined time period. Upon the completion of the processing step, the wafers are removed from the oxidation chamber, the produced oxide thickness is measured and the sequence is repeated for the next batch. The objective of this treatment is to grow a silicon oxide layer of prescribed thickness on the silicon which, in turn, is a critical factor in determining important device parameters such as threshold voltage. Note that although subsequent processing steps can be performed in order to compensate for small deviations form the

desired oxide thickness, the tolerance of such deviations is relatively narrow.

In order to control such a process, a simple adaptive scheme based on *fixed-point theorem* ideas was proposed in [1] yielding local exponential convergence of the oxide thickness to the desired set point. Due to the nonlinear nature of the process model, an important issue arising in this or similar applications is to ensure that the parameter estimates remain within the region where the approximate model of the process is valid. One approach to alleviate such a problem is to introduce a projection modification in the adaptive algorithm so that the parameter estimates are constrained in some compact set which, effectively, describes the parametric uncertainty in the process model. In the selection of this set it is of course advantageous to obtain the "smallest" possible one so as to avoid unnecessarily large perturbations away from the nominal point, caused by the growth of the higher order terms. This, in turn, motivates the study of set-membership estimation methods whose objective is precisely to determine a parametric uncertainty set, compatible with the available input-output (I/O) data and noise bounds. Note that such methods can be either off-line (batch) or on-line (recursive). The former can be useful in determining a gross estimate of the parametric uncertainty set based on a limited number of data while the latter can serve a "fine-tuning" purpose as more data become available.

The set-membership estimation problem has received a great deal of attention in the recent years [2]-[15]. Naturally, most of the available results assume linear-in-the-parameters (LITP) models where, under a bounded noise assumption, the convexity of the parametric uncertainty set allows for simple and recursive ellipsoidal approximations (e.g., [2, 3]). In the nonlinear (NLITP) case, however, an exact representation of the feasible solution set is in general not simple, since it may not be convex. The use of simply shaped sets, like axis-aligned boxes or ellipsoids has been proposed to approximate the feasible solu-

tion set [4, 7, 8]. In [5, 6, 9], the characterization the feasible set for the parameters was based on interval analysis, which is suitable to linear or nonlinear systems. These algorithms, however, rely on the batch processing of all available data and are not suitable for on-line updates of the parametric uncertainty set.

In this paper we consider the parameter estimation problem for NLITP models. We design a recursive set-membership estimator by linearizing the model of the process (with respect to the parameters) while the contribution of the higher order terms together with the possible measurement noise is viewed as the effective perturbation. Assuming the a priori knowledge of an interval containing the eigenvalues of the Hessian, the set containing the feasible parameters can be approximated by the intersection of the current parametric uncertainty set and the exteriors of two spheres. Further, using a two-step optimization scheme, we obtain a two-half-space intersection containing the feasible parameter set, for which the usual ellipsoidal approximations are applicable. It should be noted that in this approach, the bound on the effective perturbation is not constant but it decreases with a reduction of the size of the parametric uncertainty set. However, such a reduction may not be possible if the interval containing the eigenvalues of the Hessian is large relative to the size of the initial parametric uncertainty. Consequently, the applicability of the method is limited to *weakly NLITP* models, that is models for which, in the given range of operating conditions and parametric uncertainty, the contribution of the higher-order terms is sufficiently small.

Finally, to illustrate the use of this set-membership estimator, we consider the adaptive control problem of an oxidation process. For this application and following the approach of [16], the updated parametric uncertainty sets are employed as adaptation constraints, in an effort to reduce unnecessary excursions of the adaptive controller parameters.

2 Set-Membership Estimation for Weakly NLITP Models

Consider the model

$$y_k = f(x_k, \theta_*) + \epsilon_k \qquad (1)$$

where $f \in C^2$, y_k, x_k and ϵ_k are bounded sequences, $x_k \in \mathcal{X} \subset \mathbf{R}^n$, $y_k \in \mathcal{Y} \subset \mathbf{R}$, $\theta_* \in \mathcal{M} \subset \mathbf{R}^m$ is the unknown constant parameter vector, $\mathcal{M}$ is a convex bounded set (typically an ellipsoid or an intersection of a finite number of ellipsoids) and ϵ_k is a perturbation satisfying

$$|\epsilon_k| \le \mu_k \qquad (2)$$

where μ_k is an a priori known, bounded sequence. Given a measurement (x_k, y_k), our objective is to update the parametric uncertainty set $\mathcal{M}$ in a manner consistent with the noise bound μ_k while reducing its volume as much as possible. Furthermore, in order to

preserve the simplicity of its description, we restrict our attention to the case where the updated set is an ellipsoid or an intersection of a fixed number of ellipsoids.

2.1 LITP models

When the model (1) is LITP,[1] every I/O mesurement (x_k, y_k) provides a constraint for the feasible parameters, namely

$$|w_k^\mathsf{T}(\theta - \theta_*)| \le \mu_k \qquad (3)$$

where $w_k = \frac{\partial f}{\partial \theta}(x_k)$. In other words, θ_* is contained in the intersection of two half-spaces,

$$H_k \triangleq \{\theta \in \mathbf{R}^m : |y_k - w_k^\mathsf{T}\theta^\mathsf{T} x_k| \le \mu_k\} \qquad (4)$$

Thus, given n pairs of (x_k, y_k), the smallest set guaranteed to contain θ_* is $S_n = \cap_{i=1}^n H_i$. A suboptimal ellipsoid approximation of S_n can easily be obtained recursively (e.g., see [2]), starting with an initial guess $E_0 \subseteq \mathbf{R}^m$ (typically, $E_0 \supseteq \mathcal{M}$) that contains the a priori admissible values of the model parameter vector θ_*. After each successive I/O measurement (w_k, y_k) is acquired, an ellipsoid is constructed such that $E_{k-1} \cap H_k \subseteq E_k$ and such that a measure of the size of E_k is minimized. Using the volume of the set as such a measure and the notation

$$E_k \triangleq E(P_k, C_k) = \{\theta \in \mathbf{R}^m : \|(\theta - C_k)\|_{P_k}^2 \le 1\} \qquad (5)$$

where P_k is a positive definite symmetric matrix, the update equations become

$$
\begin{aligned}
e_k &= y_k - C_{k-1}^\mathsf{T} w_k \\
G_k &= w_k^\mathsf{T} P_{k-1}^{-1} w_k \\
a_k &= 1 + q(\mu_k^2 - e_k^2) + q^2 \frac{e_k^2 G_k}{1 + qG_k} \\
P_k^{-1} &= a_k P_{k-1}^{-1} - qa_k \frac{P_{k-1}^{-1} w_k w_k^\mathsf{T} P_{k-1}^{-1}}{1 + qG_k} \\
det P_k &= \frac{a_k^m}{1 + qG_k} det P_{k-1} \\
P_k &= [P_{k-1} + qw_k w_k^\mathsf{T}]/a_k \\
C_k &= C_{k-1} + qP_{k-1}^{-1} \frac{e_k w_k}{1 + qG_k}
\end{aligned}
$$

where q is a scalar parameter, selected to minimize the volume of E_k via $(m > 1)$[2]

$$
q_{opt} = \begin{cases} 0 & \text{if } \beta_2^2 - 4\beta_1\beta_3 < 0 \\ \max(0, \frac{-\beta_2 + \sqrt{\beta_2^2 - 4\beta_1\beta_2}}{2\beta_1}) & \text{otherwise} \end{cases}
$$

$$(6)$$

---

[1]That is, $f(x, \theta) = w^\mathsf{T}(x)\theta$; note that affine models can also be admitted here with some slight modifications.

[2]The boundedness of P_k, P_k^{-1} can easily be ensured regardless of the excitation properties of w_k with a small modification of the same equations [16].

$$\beta_1 = (m-1)\mu_k^2 G_k^2$$

$$\beta_2 = (2m-1)G_k\mu_k^2 - G_k^2 + e_k^2 G_k$$

$$\beta_3 = m(\mu_k^2 - e_k^2) - G_k^2$$

Note that a necessary and sufficient condition for $Vol(E_k) < Vol(E_{k-1})$ is $\beta_3 < 0$.

2.2   Weakly NLITP models

For NLITP models of satisfying (1) and (2), it is possible to employ the same approach as in the linear case by simply considering the contribution of the higher order terms as a bounded perturbation. The derivation of their bound, however, can be quite conservative if the directionality properties of the effective regressor vector $\partial f/\partial\theta$ are not taken into account. Thus, in order to incorporate some structural information regarding the higher order terms in the estimation algorithm let us suppose that:

2.1 Assumption: *For all* $\theta \in E_0$ *and all* $(x_k, y_k) \in \mathcal{X} \times \mathcal{Y}$, *the Hessian of* f *with respect to* θ *satisfies:*

$$\lambda I \leq \frac{\partial^2 f}{\partial\theta^2} \leq \sigma I \qquad (7)$$

Note that the constants σ, λ can be evaluated off-line for a given set E_0 and all possible operating points.

Given an I/O pair (x_k, y_k), a constraint on the possible values of θ_* is given by

$$\theta_* \in H_{k*} = \{\theta \in E_{k-1} : |y_k - f(x_k, \theta)| \leq \mu_k\} \quad (8)$$

In order to compute an estimate of this set, say H_k ($H_{k*} \subseteq H_k$), let θ_k' be a parameter vector such that

$$y_k = f(x_k, \theta_k') \qquad (9)$$

Assuming that $\partial f/\partial\theta \neq 0$, such a point can be found by a gradient search starting from the current estimate of θ_*, which is at least locally convergent. (With a minor modification, the same ideas can be extended to account for the case where $\partial f/\partial\theta$ may be arbitrarily small, by finding θ_k' such that $|y_k - f(x_k, \theta_k')| \leq \mu_k$.)

By taking a Taylor expansion around θ_k' we have
$$f(x_k, \theta) = f(x_k, \theta_k') + \frac{\partial f}{\partial\theta}(x_k, \theta_k')(\theta - \theta_k') +$$

$$\frac{1}{2}(\theta - \theta_k')^\top \frac{\partial^2 f}{\partial\theta^2}(x_k, \xi_k)(\theta - \theta_k') \qquad (10)$$

where ξ_k is a convex combination of θ and θ_k'. Letting $f_{\theta_k'} \triangleq \frac{\partial f}{\partial\theta}(x_k, \theta_k')$ and invoking Assumption 2.1, we arrive at the following constraints for θ_*:
$$\theta_* \in \{\theta \in E_{k-1} :$$

$$\left\{ \begin{array}{l} \sigma\|(\theta - \theta_k + \frac{f_{\theta_k'}}{\sigma})\|^2 \geq \frac{\|f_{\theta_k'}\|^2}{\sigma} - 2\mu_k \\ \lambda\|(\theta - \theta_k + \frac{f_{\theta_k'}}{\lambda})\|^2 \leq \frac{\|f_{\theta_k'}\|^2}{\lambda} + 2\mu_k \end{array} \right\} \} \quad (11)$$

Thus, using the notation $c_k^\sigma \triangleq \theta_k' - \frac{f_{\theta_k'}}{\sigma}$, $r_k^\sigma \triangleq \frac{\|f_{\theta_k'}\|^2}{\sigma^2} - \frac{2\mu_k}{\sigma}$, $c_k^\lambda \triangleq \theta_k' - \frac{f_{\theta_k'}}{\lambda}$, $r_k^\lambda \triangleq \frac{\|f_{\theta_k'}\|^2}{\lambda^2} + \frac{2\mu_k}{\lambda}$, we obtain the following estimate of H_{k*}:
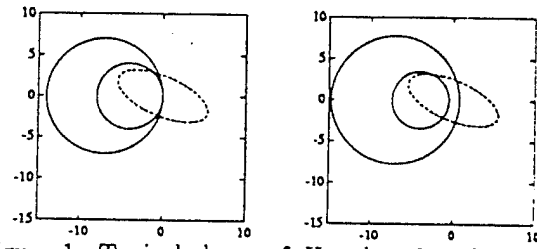
Figure 1: Typical shape of H_k when $0 \leq \lambda \leq \sigma$ or $\lambda \leq \sigma \leq 0$. ($E_{k-1}$ is denoted by the dashed line.)
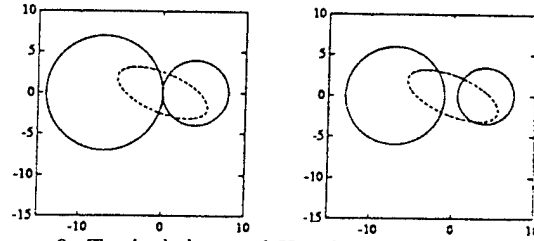
Figure 2: Typical shape of H_k when $\lambda \leq 0 \leq \sigma$. ($E_{k-1}$ is denoted by the dashed line.)

$$H_k \triangleq \{\theta \in E_{k-1} :$$

$$\sigma\|\theta - c_k^\sigma\|^2 \geq \sigma r_k^\sigma; \lambda\|\theta - c_k^\lambda\|^2 \leq \lambda r_k^\lambda\} \qquad (12)$$

Depending on the signs of λ and σ, the set H_k is the intersection of E_{k-1} and either the complement of two balls or a ball and the complement of another ball. Needless to say, in all cases H_k is a non-convex set. The possible shapes of the set H_k ($m = 2$) are illustrated in Fig.1 and Fig.2 for the cases where $\mu_k = 0$ and $\mu_k > 0$.

Our next objective is to find the smallest ellipsoid containing the intersection $E_{k-1} \cap H_k$. In order to achieve this objective, at least suboptimally, we first obtain two half-spaces whose intersection contains $E_{k-1} \cap H_k$ as tightly as possible and then use the ellipsoid update algorithm given for the case of LITP models. A relatively straightforward implementation of this idea is outlined by the following algorithm and illustrated in Fig. 3.

Algorithm

1. Find a hyperplane $L_1 = \{\theta \in \mathbb{R}^m : v_1^\top\theta = a_1\}$
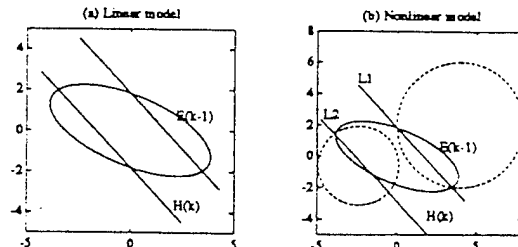
Figure 3: An illustration of the sets involved in the updates of the set E_k.

such that L_1 has at least one point on the the intersection of the boundary of the sphere (r_k^σ, c_k^σ) and the boundary of E_{k-1} and $E_{k-1} \cap H_k$ is contained in the half-space $L_1 = \{\theta \in \mathbf{R}^m : v_1^T \theta \leq a_1\}$ as tightly as possible.

2. Find a hyperplane L_2 parallel to L_1, such that L_2 has at least one point on the the intersection of the boundary of the sphere $(r_k^\lambda, c_k^\lambda)$ and the boundary of E_{k-1} and such that L_1 and L_2 enclose tightly the intersection $E_{k-1} \cap H_k$

3. Using L_1 and L_2 in the equations for the LITP-model case, update the ellipsoid E_{k-1}.

4. Repeat steps 1-3 by using the sphere $(r_k^\lambda, c_k^\lambda)$ to define L_1 and the sphere (r_k^σ, c_k^σ) for L_2 until convergence is achieved (typically, a couple of iterations are sufficient).

An important problem arising in the construction of the above algorithm is that the computation of the hyperplane parameters v_1, a_1 should be as optimal as possible under the constraint that the the intersection $E_{k-1} \cap H_k$ should be entirely contained in the associated half-space. Note that, although a suboptimal computation may result in a deterioration of the speed of convergence and a larger set E_k, a violation of the constraint may exclude feasible points from E_k. Since the latter would have severe consequences in the intended application of the algorithm, the computations should be performed so as to *guarantee* that under no circumstances the above constraint is violated. For this purpose, we adopt the following two-step procedure for the computation of the parameters defining the hyperplance L_1:

1. Given a direction v, find the smallest a such that $E_{k-1} \cap H_k$ is contained in the half-space $v^T(\theta - \theta_k') \leq a$.

2. Use any optimization scheme to minimize a with respect to v. (A typical starting point is the vector $c_k^\sigma - \theta_k'$.)

It is straightforward to show that the first step of the above procedure reduces to the problem

$$\max_y (y - c)^T (y - c)$$
$$\text{s.t. } y^T P y = 1 ; \quad P = P^T > 0$$

where $y, c \in \mathbf{R}^{m-1}$. Using a Lagrange multiplier Λ, the solution of this problem can be obtained iteratively, via a constrained Newton algorithm, by finding the unique Λ satisfying $(I + \Lambda P)y = c$ and $\Lambda < -1/d_{min}$ where d_{min} denotes the minimum eigenvalue of P. Note that the convergence to the desired solution, within any desired tolerance, is guaranteed provided that the Newton iterations are constrained in the interval $(-\infty, -1/d_{min})$ (since other local extrema exist outside this interval) and that c is not an eigenvector of P (this can always be achieved by a small
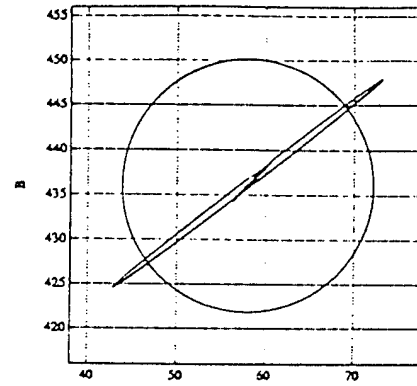
Figure 4: Recursive computation of the parametric uncertainty sets for the silicon oxidation example.

perturbation of c). Finally, regarding the second step, although convergence to local minima may be possible, the computation of the objective function 'a' via the above procedure ensures that even a suboptimal computation preserves the property $\theta_* \in E_k$.

2.3 An example of recursive Set-Membership estimation for weakly NLITP models

To demonstrate the behavior of the above algorithm, let us consider the following simplified dynamic model of a silicon oxidation process [17]

$$\dot{x} = \frac{B}{2x + A} \tag{13}$$

where x is the oxide thickness, and A, B are experimentally determined constants. The solution of this differential equation is

$$t = \frac{x^2 - x_0^2}{B} + \frac{A(x - x_0)}{B} \tag{14}$$

Letting $x_0 = 70$ and given that $A \in [48, 68], B \in [426, 446]$, $x \in [100, 500]$, $\mu_k = 0.01$ we would like to recursively estimate a set containing the "actual" parameters, A_*, B_* based on the following measurements

Table 1

k	1	2	3	4	5
x_k	438.39	264.83	436.60	207.73	266.16
t_k	461.23	168.11	457.51	101.19	169.83

6	7	3	9	10
314.92	287.17	214.88	171.33	161.49
238.84	198.26	108.79	66.19	57.74

(Here we used $A_* = 58, B_* = 436$.)

For this problem, the algorithm outlined in the previous subsection, with the Hessian bounds being

1069

$\sigma = 0.00782$ and $\lambda = -0.00074$, yields the parametric uncertainty set estimates shown in Fig. 4. Note that the volume of this set is reduced rapidly to a rather small value ($\det P_0^{-1} = 40000, \det P_1^{-1} = 65.57, \det P_2^{-1} = 0.125, \dots \det P_{10}^{-1} = 0.125$).

3 Application to the adaptive control of a silicon oxidation process

In this section we consider the batch-to-batch adaptive control problem of a silicon oxidation process where the objective is to adjust the oxidation time so that the thickness of the produced oxide converges to a desired value, denoted by x_* [1]. We assume that the actual process is governed by the following differential equation

$$\dot{x} = \frac{B}{2x + A} + Ce^{-x/L} \qquad (15)$$

which accounts for an initial reaction-dominated stage [17]. However, to emulate modelling inaccuracies, all the derivations associated with the controller design and operation are based on the simplified model

$$t = \frac{x^2 - x_0^2}{B} + \frac{A(x - x_0)}{B} \qquad (16)$$

Following [1], a simple adaptive scheme can achieve this objective, at least locally. During the process, however, it is likely that the lack of sufficient excitation for identification purposes may cause the parameter estimates to converge to physically unacceptable values (e.g., depending on the initial conditions A may converge to a negative value). Although in this particular case, such an event does not affect the computation and convergence of the control input, it is desirable to implement constraints in the adaptation so that the parameter estimates are within physically acceptable bounds.[3] Further, since the convergence of the adaptation is only local, it is also desirable to restrict the parameter estimates in a small compact set that is reduced on-line as new measurements are obtained.

To illustrate the the implementation of these ideas via on-line Set-Membership estimation, we employ the following adaptive controller for the silicon oxidation process:[4]

$$t_k = f(x_*, \theta_k)$$

---

[3] Such considerations may become important in more general cases where the process model is of the implicit form $F(x, u, \theta) = 0$; in such cases, the ability to compute an appropriate control input u may depend critically on the value of θ.

[4] This is a slightly different (simpler) version of the algorithm used in [1]; both algorithms, however, have similar qualitative properties and similar parameter projection techniques are applicable to both cases.
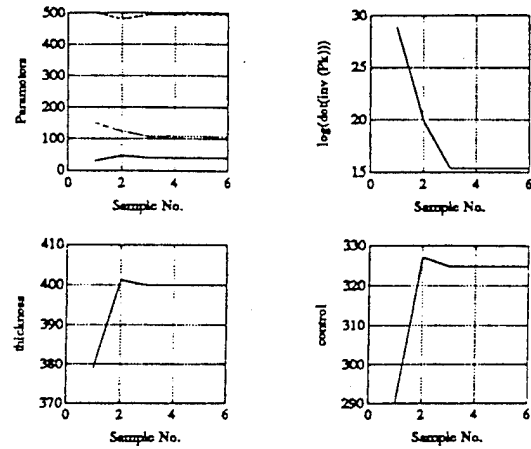
Figure 5: Simulated response of the adaptive controller; $\mu_k = 4.2$

$$\theta_{k+1}^1 = \theta_k \frac{\partial f}{\partial \theta}^\dagger (x_k, \theta_k)[t_k - f(x_k, \theta_k)]$$

$$\theta_{k+1} = \mathcal{P}_{\mathcal{M} \cap E_k \cap H_k^L}(\theta_{k+1}^1)$$

where $z^\dagger = z^T[zz^T]^{-1}$ for a row vector z, $\mathcal{P}_X$ is a projection on the convex set X [18].[5] Further, E_k denotes the updated parametric uncertainty set and H_k^L is the intersection of the two half-spaces containing $E_k \cap H_k$. Both of these sets are computed by the Set-Membership algorithm of the previous section, operating in parallel with the adaptive controller.

The properties of the above adaptive controller are similar to that in [1]. That is, locally around (x_*, θ_*)

1. $\theta_k \in \mathcal{M} \cap E_{k-1}, \forall k$.

2. The tracking error $e_k = x_k - x_*$ converges to the residual set $\{e : e \leq d(\sigma, \lambda, \mu_k)\}$ exponentially fast, where d is a dead-zone-like threshold s.t. $d \to 0$ as its arguments approach zero.

3. In the ideal case where $\mu_k = 0$, the tracking error e_k converges to zero, even if $d > 0$.

Applying this adaptive controller to the silicon oxidation example, with

$$f(x, \theta) = \frac{x^2 - \theta^2(3) + \theta(1)(x - \theta(3))}{\theta(2)}$$

where $\theta = [A, B, x_0]^T$. The results of this example are as shown in Figures 5 and 6.

For the simulations, we used the following nominal values: $A = 3854.87 exp(-5336.43/T)$; $B = 2.8333E11 exp(-25754.06/T)$; $C = 20\frac{2.8333E11}{3854.87} exp(-25754.06/T)$; $L = 100\text{Å}$; where T is in $°K$, B in $\text{Å}^2/s$ and A in $\text{Å}/s$. In all cases,

---

[5] When X is an intersection of sets X_i, the projection is computed by sequential orthogonal projections on each X_i until convergence is achieved.

Figure 6: Simulated response of the adaptive controller with a more conservative selection of the perturbation bounds; $\mu_t = 10$

the temperature initially follows a ramp with slope $300^\circ C/s$ and after reaching $1000^\circ C$, remains constant for the rest of the operation.

The desired oxide thickness is $x_* = 400\text{Å}$ while the region of operation is between 200Å to 600Å. A nominal value for the parameter θ was computed off-line (e.g., in practice by processing a f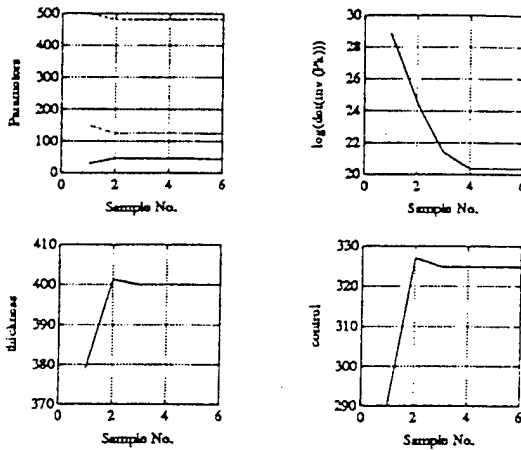ew test wafers) as $\theta_o = [67.6, 513.8, 100.8]$. This value was obtained by using a nonlinear least squares fit, with the constraints: $A \in [0, 95]$, $B \in [450, 538]$, $x_0 \in [0, 205]$. The corresponding parametric uncertainty set is $\theta(1) \in [47, 71]$, $\theta(2) \in [15, 1215]$, $\theta(3) \in [0, 334]$. The bound of the measurement noise in terms of the oxidation time is 4.2 seconds while $\sigma = 0.0096$ and $\lambda = -0.0060$. Finally, for the initialization of the adaptation we selected $\theta_0 = [30, 500, 150]^T$, $E_0 = \{\theta \in \mathbb{R}^m : (\theta - c_0)^T P_0^{-1} (\theta - c_0) \leq 1\}$ where $c_0 = [47.5, 494, 167]^T$ and

$$P_0 = \begin{pmatrix} 0.1123 \times 10^{-3} & 0 & 0 \\ 0 & 0.1860 \times 10^{-3} & 0 \\ 0 & 0 & 0.0139 \times 10^{-3} \end{pmatrix}$$

whose inverse has determinant $det(P_0^{-1}) = 3.4538 \times 10^{12}$.

As demonstrated by the simulations, the convergence of the algorithm is fairly rapid while the parameters remain within physically meaningful bounds. Furthermore, although the obtained measurements do not provide sufficient (or even persistent) excitation for identification purposes, the volume of the parametric uncertainty set is reduced considerably.

References

[1] K. S. Tsakalis and P. E. Crouch, "A Simple Adaptive Controller for an Oxidation Process," *Proc. 1992 ACC*.

[2] E. Fogel and Y. F. Huang, "On the Value of Information in System Identification - Bounded Noise Case," *Automatica*, Vol. 18, No. 2, pp.229-238, 1982.

[3] J. P. Norton, "Identification and application of bounded-parameter models," *Automatica*, 23, 497-507, 1987.

[4] M. Milanese and A. Vicino, "Estimation Theory for Non-liner Models and Set Membership Uncertainty," *Automatica*, Vol. 27. No.2, pp. 403-408, 1991.

[5] L. Jaulin and E. Walter, "Set Inversion via Interval Analysis for Nonlinear Bounded-Error Estimation," Automatica, Vol. 29, NO.4 pp. 1053-1064, 1993.

[6] L. Jaulin and E. Walter, "Guaranteed nonlinear parameter estimation from bounded-error data via interval analysis," *Math. Comput. Simulation*, 35, 123-137, 1993.

[7] G. Belforte and M. Milanese, "Uncertainty intervals evaluation in presence of unknown but bounded errors, Nonliear families of models," *Proc. 1st IASTED Symp. on Modelling, Identification and Control*, Davos, pp. 75-79, Feb. 1981.

[8] M. Milanese and G. Belforte, "Estimation theory and uncertainty interval evaluation in the presence of unknown but bounded errors: linear families of models and estimators." *IEEE Trans. Automat. Contr.*, Vol. AC-27, pp.408-414, Apr. 1982.

[9] R. E. Moore, "Parameter sets for bounded-error data," *Math. Comput. Simulation*, 34 (2), 113-119, 1992.

[10] E. Walter and H. Piet-Lahanier, "Robust nonlinear parameter estimation in the bounded noise case," *Proc. 25th IEEE CDC*, Athens, Greece, pp. 1037-1042, Dec 1986.

[11] T. Clement and S. Gentil, "Reformulation of parameter identification with unknown but bounded errors," *Math. Comput. Simulation*, 30, 257-270, 1988.

[12] G. Belforte, B. Bona and V. Cerone, "Parameter Estimation Algorithms for a Set-membership Description of Uncertainty," *Automatica*, Vol. 26, No. 5, pp. 887-898, 1990.

[13] S. H. Mo and J. P. Norton, "Fast and robust algorithm to compute exact polytope parameter bounds," *Math. Comput. Simulation*, 32 (5 and 6), 481-493, 1990.

[14] E. Walter and H. Piet-Lahanier, "Exact Recursive Polyhedral Description of the Feasible Parameter Set for Bounded-Error Models," *IEEE Trans. Automat. Contr.*, Vol. 34, No. 8, Aug. 1989.

[15] J. P. Norton, "Identification of parameter bounds for ARMAX models from records with bounded noise," *Int. J. Control*, Vol. 45, 375-390, 1987.

[16] K. S. Tsakalis and S. Limanond, "Asymtotic Performance Guarantees in Adaptive Control," *Int. J. Adapt. Contr. Sig. Proc.*, to appear, 1994.

[17] H. Massoud and J. Plummer, "Thermal Oxidation of Silicon in Dry Oxygen," *J. Electrochem. Soc., Solid State Sci. Tech.*, July 1985.

[18] G.C. Goodwin and K.S. Sin, *Adaptive Filtering Prediction and Control*. Prentice Hall, 1984.

[19] V.M. Kuntzevich and M. Lychak, *Guaranteed Estimates, Adaptation and Robustness in Control System*. Springer-VerlagBerlin Heidelberg, 1992.

A Simple Adaptive Optimization Algorithm for the Tungsten LPCVD process *

Timothy S. Cale[†], Peter E. Crouch[‡], Sisan Shen[‡], and Kostas S. Tsakalis[‡]

[†]Center for Solid State Electronics Research

[‡] Center for System Science and Engineering

Arizona State University.

Box 877606, Tempe, AZ 85287-7606

c/o Peter.Crouch@asu.edu

Abstract

In this note, a reduced order, physically-motivated empirical model is proposed and validated via simulation for the single wafer tungsten Low Pressure Chemical Vapor Deposition (LPCVD) processing step. The so-called multiple response surface method is adopted to describe the spatial deposition nonuniformity across a wafer surface. Based on this modeling methodology, a simple adaptive optimization control strategy is developed by which the average deposition thickness at the wafer surface is controlled to a desired level while its variation of the state across the wafer surface is minimized. Simulation results demonstrate the effectiveness of the control strategy and its potential capability of rejecting disturbances during the process. In this study, a simulation platform (CFDSWR) is used to represent the single wafer tungsten LPCVD process. The control strategy introduced here is quite general and applicable to other processing steps as well.

1   Introduction

In the semiconductor manufacturing industry, the process by which an integrated circuit (IC) is fabricated involves a large number of processing steps. Each processing step involves a single wafer, or a batch of wafers in a boat, being placed in various pieces of equipment to accomplish tasks such as oxidation, reactive ion etch, chemical vapor deposition, lithography, etc. Many of these pieces of equipment are basically reactor vessels into which the wafer (or wafers) is placed and then exposed to controlled flows of gases and reactants that change the characteristics of the surface layers of the wafer through either a chemical or a physical process, or both. Each processing step is controlled by maintaining the manipulated variables (control inputs) near appropriately selected set-points, e.g., temperature, pressure, gas flows and processing time.

The first difficulty associated with this control problem is the modeling of the process. A considerable

amount of effort has been devoted to the mathematical modeling of many individual processes based on the underlying physical and chemical mechanisms [1] [2]. Unfortunately, almost none of these models is suitable for control purposes, mainly due to their complexity and the large number of parameters that must be determined. In an attempt to alleviate these problems, polynomial models are frequently used to approximate the static state response of a process step. The parameters of the model are determined by applying traditional response surface techniques and linear regression schemes [3]. However, the practical usefulness of the polynomial model is limited by the complexity often required to yield the desired accuracy over the range of operating conditions. In turn, such models may be over-parametrized for a smaller range of operating conditions, something that may cause parameter convergence/drift problems when on-line adaptation is used in the presence of disturbances [4]. Although a remedy for these problems can be found in the use of adaptation dead-zones, it is reasonable to seek an empirical, but physically motivated model whenever possible, in an effort to minimize the number of adjustable parameters. This is the approach adopted in this study.

Even though set-points of the manipulated variables for a process may be kept constant, process conditions and the associated measurements generally drift or fluctuate during the machine operation. It is therefore important to detect these trends in the process and adjust the set-points to compensate for their effects. This is both a real time problem, occurring during the processing of a wafer (or batch of wafers) using in-situ measurements, and a run-to-run problem in which machine settings are adjusted between runs. To alleviate these problems, several control strategies have been proposed including statistical process control [5], model based control [6] and model prediction and correction [7], etc. For the same purpose, two of the authors (Tsakalis and Crouch) have proposed an adaptive control scheme [8] which has the potential of addressing this issue in the case of run-to-run control. An application of this technique has been considered in the experimental manufacturing process of a capacitor

Figure 1: Schematic diagram of the Single Wafer LPCVD reactor

Figure 2: Relative error between simulation and prediction data of deposition rate
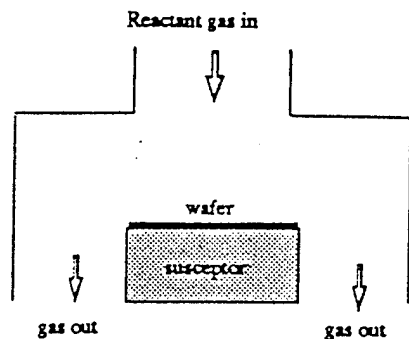
[9]. Suitable variations are concurrently being developed to deal with real time issue. As the wafer size used in device fabrication continues to increase, it becomes more difficult to maintain acceptable processing uniformity across the wafers. Consequently, not only must the average processing condition on the wafer surface be accurately controlled, but it is also equally important to minimize the variation of the processing conditions across the wafer surface at the same time. Regarding the latter issue alone, some experimental and analytical (or numerical) methods have been investigated (see, e.g., [10]). These studies are mainly concerned with the optimal design of the processing reactor vessel or equipment operation set points, or both. More recently some control problems has been considered for Rapid Thermal Processing (RTP) processes where the distribution of power to the heater lamps was controlled so as to achieve the most uniform temperature distribution possible on the wafer surface and, hence, indirectly improve the thickness uniformity [7] [11]. In general, little effort has been devoted to the combined issue of accurately controlling the average processing conditions while minimizing their variation. These are precisely the main objectives of the present study.

2 Model development for single wafer tungsten LPCVD

Chemical Vapor Deposition (CVD) is one of the important processing steps in IC fabrication; it uses volatile compounds, containing the species to be deposited, to form a thin solid film on a solid surface such as a silicon wafer. Single wafer Low Pressure CVD (LPCVD) is becoming more competitive compared with its multi-wafer counterpart due to its great potential in improving wafer-to-wafer deposition uniformity. A schematic diagram of a typical single wafer LPCVD reactor is illustrated in Fig.1.

In the tungsten LPCVD process, a thin layer of solid tungsten is to be deposited on a wafer surface, with tungsten hexafluoride (WF_6) and hydrogen (H_2) as the reactant gases. The chemistry governing the process is
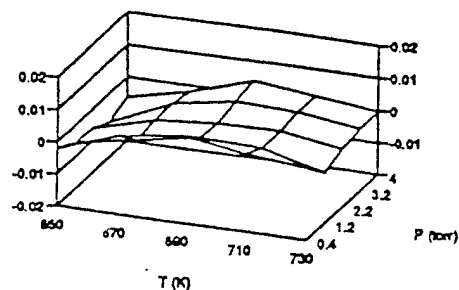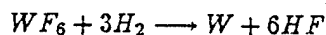
$$WF_6 + 3H_2 \longrightarrow W + 6HF$$

The manipulated variables (control inputs) involved in the processing step are the susceptor temperature T (K), total pressure P ($torr$) and flow rates of WF_6, H_2 and inert carrier gas ($sccs$). The output variables (controlled outputs) are deposition thickness (or growth rate) and its spatial nonuniformity across the wafer surface. For control purposes, the modeling of such a process aims to establish relationships between the manipulated and output variables. Before any candidate models are proposed, experimental design and statistical analysis are usually helpful in determining a suitable model structure. In this note, we discuss the use of a simulation test bed rather than an actual reactor for model development. Simulation data are generated using a simulation platform called CFDSWR [16] [17], that simulates an LPCVD reactor. A full factorial experiment design was adopted and Yate's analysis [12] was applied to the simulation data in order to determine the most significant relationships between the manipulated variables and the output variables. The results show that the susceptor temperature T and the total pressure P are the two dominant factors which affect both deposition growth rate and its nonuniformity. Based on a physical analysis given by [13] and in combination with the above statistical analysis results, we developed a reduced-order model relating the tungsten deposition growth rate (GR) with the susceptor temperature and the total pressure at a certain point on the wafer surface. This reduced-order model is given as

$$GR = K_0 e^{(\theta_1/T)} \frac{\theta_2 P^{3/2}}{1 + \theta_3 P} \qquad (1)$$

where K_0 is a constant, θ_1, θ_2 and θ_3 are unknown model parameters determined by fitting the model to the simulation data (using, for example, a Gauss-Newton nonlinear regression algorithm [14]).

Modeling the spatial nonuniformity normally presents a challenge since the underlying map from the manipulated variables to the nonuniformity is in general very complicated. A possible approach is to include spatial information into the model parameters introduced above so that the model may be capable of "reflecting" the spatial variations of the deposition and hence the nonuniformity can be defined based on the model. On the other hand, the so-called multiple surface response method, developed by [15], successfully circumvents the difficulty by using a simpler

strategy. Since in most cases, the spatial nonuniformity is not measured directly but is calculated from measurements of other output characteristics at specified locations, it may be more efficient to fit the measurements first and then construct the model for the spatial nonuniformity. This is different from the traditional approach of first calculating the value of the nonuniformity based on the measurements and then fitting the model parameters accordingly. For example, given N deposition thickness measurements at different points along the wafer radius and assuming the processing conditions have a circular symmetry, a suitable measure of the deposition nonuniformity is the standard deviation divided by the average of these thickness measurements. Fig.2 shows the error between the simulation data and the model predicted ones. The proposed model is able to accurately predict the deposition growth rate over a fairly broad range of both temperature and pressure.

3 Adaptive Control of the Deposition Process

Let us assume that a processing step can be approximated by a model

$$\hat{x} = f(u, \theta) \tag{2}$$

where $\hat{x}$ denotes the predicted state (vector) starting from an initial state x_0 after a certain amount of processing time. u stands for the set point of the manipulated variables. Note that the processing time can also be considered as a control input, included in u. Finally, θ denotes a vector of unknown, adjustable parameters. We assume that there exists a vector θ^* for which (2) is a "good" approximation of the actual process; i.e., $x = f(u, \theta^*)$, x is a measured state which is available from the processing step. Our first objective is to control the state x to a desired level x^* so that a certain prescribed value y^* depending on x^* is achieved. That is,

$$y^* = \Phi(x^*) = g(u, \theta^*) \tag{3}$$

The second control objective can be described as

$$\min_u J(x, u) = \min_u h(u, \theta^*) \tag{4}$$

where J denotes any suitable cost functional, defined in terms of the state and control input. If θ^* were known, then (3) and (4) could be used to determine desirable, locally unique, values of the control inputs. However, in actual control problems, the value θ^* is partially unknown. Hence, we seek an iterative adaptive control algorithm enabling the computation of updates of the parameter and control vectors, while asymptotically achieving (at least approximately) the above two objectives. Conceptually, the adaptation must be driven by a computable error $(\hat{x} - x)$, together with the optimization problem

$$\min_u h(u, \theta) \tag{5}$$
$$s.t. \; g(u, \theta) - y^* = G(u, \theta) = 0 \tag{6}$$

To generate a candidate algorithm, we re-cast the problem as the solution of an underdetermined system of nonlinear equations. We let λ denote the Lagrange multiplier and set

$$H(u, \theta, \lambda) = \frac{\partial [h(u, \theta) + \lambda G(u, \theta)]^T}{\partial u} \tag{7}$$

Then the control problem may be reformulated as one of finding solutions of the following system of nonlinear equations

$$\begin{cases} f(u, \theta) - f(u, \theta^*) = 0 \\ H(u, \theta, \lambda) = 0 \\ G(u, \theta) = 0 \end{cases} \tag{8}$$

with respect to the vectors θ, u and λ. This is basically the approach adopted in the paper by two of the authors (Tsakalis and Crouch) [8]. The idea employed there is to generate run-to-run adaptive control algorithm by solving the equations iteratively using a generalized form of Newton's algorithm. Although the same method may also be applied in our case, the possibly high sensitivity of u^* (optimal solution of u) on the parameters θ and the coupling of the parameter and control update rules, often results in demanding unacceptably large variations of the control input. To circumvent this problem, one could employ parameter projection techniques [4, 18], possibly combined with modifications of the cost functional so as to penalize excessive control inputs.

In this study, however, we adopt a simpler alternative approach (motivated by the indirect adaptive control paradigm) in which the parameter update and control update are decoupled. That is, we first compute the parameter update by applying a sequence of nonlinear regression steps to the first components of the equation (8)

$$\theta(k + 1) = \theta(k) + \frac{\partial f^\dagger(u(k), \theta(k))}{\partial \theta}(\hat{x}(k) - x(k)) \tag{9}$$

until $|(\hat{x}(k) - x(k))/x(k)| \leq \epsilon$ where $\frac{\partial f^\dagger(u(k), \theta(k))}{\partial \theta}$ is the pseudo inverse of $\frac{\partial f(u(k), \theta(k))}{\partial \theta}$ and then update the control $u(k + 1)$ using (5) and (6) with θ replaced by $\theta(k + 1)$. In general, we cannot expect $\theta(k)$ and $u(k)$ to converge to θ^* and u^*. It is therefore unreasonable to expect that the optimal value of $h(u^*, \theta^*)$ of the minimization problem is achieved asymptotically, even in the ideal case of exact models. However, in view of the available indirect adaptive control results, we may argue that such an adaptive control algorithm converges to a residual set ($\forall k \geq k_0$) where

$$\begin{aligned} x(k) &= f(u(k), \theta(k)) = f(u^*, \theta^*) = x^* \\ \Phi(x(k)) &= \Phi(x^*) = y^* \\ |h(u(k), \theta(k)) &- h(u^*, \theta^*)| \leq \sigma \end{aligned} \tag{10}$$

In our studies we used the threshold value $\epsilon = 0.001$ that resulted in $\sigma \leq 0.05$. It should be mentioned, however, that the dependence of σ on the adaptation parameters and the range of operating conditions is rather complicated; and although this dependence could be determined via a tedious worst-case analysis, such a direction was not pursued in our study.
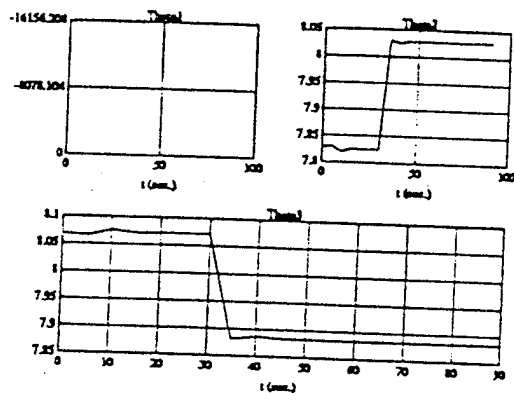
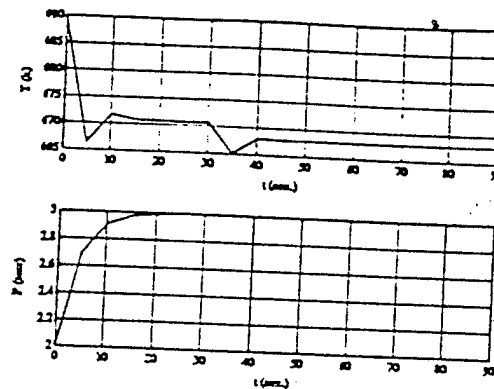Figure 3: Adaptation of the model parameters during the processes

Figure 4: Control inputs: Temperature (T) and Pressure (P)

4 Simulations and Discussion

To test the proposed strategy for controlling single wafer tungsten LPCVD processing step, the CFDSWR is used to represent the actual process. The susceptor temperature (T) and total pressure (P) are selected as the two manipulated variables while the other three variables (flow rates of reactant gases) are kept constant. The output variables are deposition thickness and its spatial nonuniformity. We assume that the thickness measurements denoted by z_i are measured at different points along the wafer radius in five-second time intervals. Four measurement points are obtained from the simulation with more samples near the edge of the wafer where the variation in process conditions is greatest. We also assume that the instantaneous deposition rate measurements at each point x_i ($\mathring{A}/min$) are not available, but can be approximated by subtracting the current measured thickness by the previously measured one and dividing by the time differences (5 sec).

Our first objective is to deposit a desired spatial average thickness ($810\mathring{A}$) of solid tungsten on the wafer surface in a given period of time (90 sec). We also want to minimize the deposition nonuniformity based on four thickness measurements. Here, we consider an operating range of $T \in [650, 730](K)$ and $P \in [1.0, 3.0]$ (torr) with flow rates of WF_6, H_2 and inert gas being $[0.25, 5.0, 8.0]$ (sccs) respectivly. The first objective can be met by following one of two approaches. The first is to control the average deposition rate to a desired value. The other is a quasi-real-time strategy in which we first pre-define a reference thickness trajectory based on the first objective and then try to control the average thickness growth to follow the reference trajectory. In our study, the second strategy is adopted due to its potential capability of rejecting disturbances and drift during processing. A constant thickness increment (per five seconds) trajectory is used in the simulation. Note that we do not imply that the constant thickness increment trajectory is the "best" one with respect to the final deposition nonuniformity. In fact determining an optimal reference trajectory is an interesting problem in itself which is left as a topic of future research.
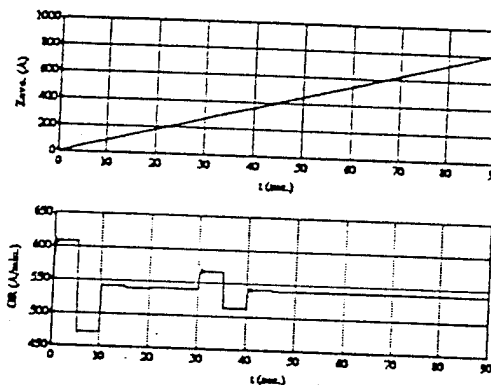
Figure 5: Thickness Trajectory Z_{ave} ($\mathring{A}$) and the corresponding Growth rate GR ($\mathring{A}/min.$)

In the test simulation, the constrained optimization problem was solved using standard numerical optimization algorithms (MATLAB). The simulation results are shown in Fig. 3–7. Fig. 3 shows the adaptation progress of the model parameters corresponding to the first measurement point. The set-point changes of the control inputs (Fig. 4) indicate that, in order to minimize the nonuniformity, we must decrease the processing temperature as much as possible. On the other hand, the pressure must be increased so as to maintain suitable deposition rate. Fig. 5 shows the average thickness growth trajectory Z_{ave} and the corresponding growth rate GR generated from the CFDSWR at 0.2 second time intervals. The actual thickness nonuniformity defined as the standard deviation divided by the average is shown in Fig. 6.

To test the capability of rejecting process disturbances, we introduce a step change increase of hydrogen flow rate at 30 seconds. After the change, the susceptor temperature decreases in response to the disturbance and quickly converges to a new level minimizing spatial nonuniformity subject to the desired average growth rate constraint. On the other hand, the pressure still remains the same in order to maintain the desired growth rate (Fig. 4). This behavior simply indicates that for the given range of operating conditions, temperature has the most dominant effect
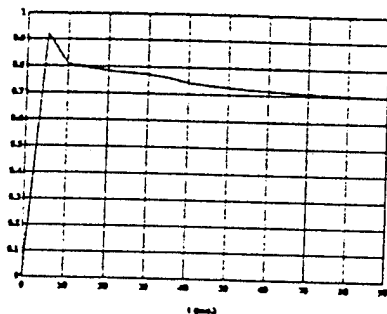
Figure 6: Thickness nonuniformity defined as the standard deviation of the 4 measurements divided by the average

Figure 7: Error between the measured and the desired thickness

on the minimization problem, with the minimum occuring at the boundary of the control space where P is at its maximum value. Fig. 7 shows the error between the desired thickness values defined by the reference trajectory and the "actual" one generated from the CFDSWR at each measurement time. The thickness deviation caused by the disturbance vanishes quickly after a few samples.

5 Conclusions

An accurate, yet simple, physically-motivated empirical model for the single wafer tungsten LPCVD processing step is considered. Adopting a multiple response surface method, a reasonably accurate description of the spatial nonuniformity is obtained. A simple adaptive optimization strategy is introduced by solving a decoupled problem of model parameter updates and control updates. The effectiveness of this control strategy is illustrated by means of a simulation example. It should be noted that in an application to an actual process, rather than a simulation platform, the temperature cannot be changed instantaneously but must be controlled by regulating the power supplied to the susceptor heater. However, since the relationship between the power and the susceptor temperature can typically be approximated by a first order ODE, such a description can easily be incorporated in our formulation, leaving the main idea of the adaptive optimization scheme largely unaffected. This, together with robustness issues arising from noisy measurements and/or modeling errors as well as the experimental validation of such control schemes are left as topics of future work.

References

[1] R. Arora, and R. Pollard, "A Mathematical Model for Chemical Vapor Deposition influenced by Surface Reaction Kinetics: Application to Low Pressure Deposition of Tungsten", J. Electrochem. Soc., 138(5), 1523 (1991).

[2] T. S. Cale, T. H. Gandy, G. B. Raupp, and M. K. Jain, " Comparison of Continuum Diffusion - Reaction and Particulate Ballistic Transport Deposition Models for LPCVD", in Tungsten and other Advanced Metals for ULSI Appli-

cations in 1990, G. C. Smith, and R. Blememthal Eds., MRS Publisher, Pittsburgh, 1991.
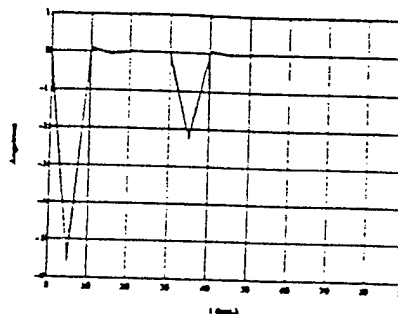
[3] K. K. Lin, J. Huang, and C. J. Spanos, "Statistical Equipment Modeling for VLSI Manufacturing", Symp. Automated Semiconductor Manufacturing, 176th Electrochemical Soc. Mtg., Aug. 1990.

[4] S. Sastry, and M. Bodson, Adaptive Control Stability, Convergence and Robustness, Prentice Hall, 1989.

[5] Sovarong Leang "Supervisory Control System for a Photolithographic Workcell", SRC Techn. Report T92-181.

[6] R. A. Soper, D. A. Mellinchamp, and D. E. Seborg, " An Adaptive Nonlinear Control Strategy for Photolithography". Proc. 1993 ACC, 2998.

[7] T. Breedijk, T. F. Edgar, and L. Trachtenberg, "A Model Predictive Controller for Multivariable Temperature Control in Rapid Thermal Processing", Proc. 1993 ACC, 2980.

[8] K. S. Tsakalis, and P. E. Crouch, "A Simple Adaptive Controller for an Oxidation Process", Proc. 1992 ACC.

[9] K. D. Stoddard, K. S. Tsakalis, P. E. Crouch, and M. Kozicki, " Application of Adaptive Feed - Forward Control to Parallel - Plate Capacitor Manufacturing", Proc. 1994 ACC.

[10] H. T. Tran, and J. S. Scroggs, "Modeling and Optimal Design of a Chemical Vapor Deposition Reactor", Proc. 31st CDC (1992).

[11] C. D. Schaper, "Real Time Control of Rapid Thermal Processing Semiconductor Manufacturing Equipment" Proc. of ACC P2985 (1993).

[12] R. E. Walpole, and R. H. Mayers, Probability and statistics for Engineer and Scientist, Macmillan Publishing Company, 1985.

[13] C. M. Mcconica, and K. Krishnamani, " The Kinetics of LPCVD Tungsten Deposition in a Single Wafer Reactor", J. Electrochem. Soc., 133 (12), 2542, 1986.

[14] D. M. Bates, and D. G. Watts, Nonlinear Regression Analysis and its Application, John Wiley and Sons, 1988.

[15] R .S. Guo, and E. Sachs, "Modeling, Optimization and Control of Spatial Uniformity in Manufacturing Processes", IEEE Trans. Semicond. Manufact., Vol.6, No.1, 41, 1993.

[16] J. H. Park, User's guide for CFDSWR, A Computational Fluid Dynamics of a Single Wafer Reactor, Arizona State Univ. 1993.

[17] J. H. Park, Simulation of Low Pressure Chemical Vapor Deposition using combined reactor scale and feature scale models, Ph.D thesis, Arizona State Univ, Aug.1992.

[18] K. S. Tsakalis, L. Song, "Set-membership estimation for weakly nonlinear models: An application to the adaptive control of semiconductor manufacturing process", Proc. 33rd CDC, 1994.

Run-to-Run Adaptive Optimization of a Tungsten Silicide LPCVD Process *

Timothy S. Cale[†], Peter E. Crouch[‡], Sisan Shen[‡], and Kostas S. Tsakalis[‡]

[†]Center for Solid State Electronics Research, [‡] Center for System Science and Engineering
Arizona State University, Tempe, AZ 85287-5706
email: tsakalis@enuxsa.eas.asu.edu, FAX: (602) 965-0461

KEYWORDS: Run-to-run Control; LPCVD processes; Adaptive control; Constrained Optimization
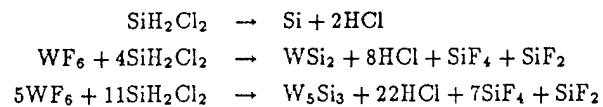
Summary

In this note we introduce a novel modeling technique, in an effort to develop a physically-motivated empirical model of the deposition rate and spatial deposition nonuniformity for a single-wafer Tungsten Silicide Low Pressure Chemical Vapor Deposition (LPCVD) processing step. In general, such a description is difficult to obtain due to the complexity of the mapping between the spatial nonuniformity and manipulated variables. Combining the so-called multiple response surface method [5] with a "feedback-like" model structure, relating reactant partial pressures with manipulated variables, we develop a reasonably accurate description of the deposition rate and spatial deposition nonuniformity across wafer surface. This model offers good fitting accuracy with fewer adjustable parameters compared with the more traditional polynomial-structure models.

Further, based on this modeling methodology, we derive a run-to-run adaptive control/optimization strategy aiming to achieve prescribed values of the average deposition rate and silicon to tungsten ratio at the wafer surface, while minimizing the variation of the deposition rate across the wafer surface. Such objectives are critical, and quite common, in the modern semiconductor industry. The effectiveness of this control strategy is demonstrated by simulation results, using the simulation platform CFDSWR [3, 4] to represent the single wafer tungsten silicide LPCVD process. (The CFDSWR is a computer program that simulates the reaction and transport phenomena occurring during LPCVD processes.)

In order to describe the basic idea behind our simplified, control-oriented modeling approach, we consider a Tungsten Silicide LPCVD process. In this process the reactants, tungsten hexafluoride (WF$_6$) and dichlorosilane (SiH$_2$Cl$_2$; DCS), are fed in a reactor containing the wafer. For the deposition of tungsten silicide (WSi$_x$) films, at least three reactions, depositing Si, WSi$_2$, and W$_5$Si$_3$, seem to contribute to the apparent deposition rate and film composition [1]:

$$SiH_2Cl_2 \rightarrow Si + 2HCl$$
$$WF_6 + 4SiH_2Cl_2 \rightarrow WSi_2 + 8HCl + SiF_4 + SiF_2$$
$$5WF_6 + 11SiH_2Cl_2 \rightarrow W_5Si_3 + 22HCl + 7SiF_4 + SiF_2$$

Although the reaction stoichiometries have not been formally established, experimental measurements have been used to correlate the apparent WSi$_x$ deposition rate and film composition with the wafer temperature and reactant partial pressures through (apparent) reaction rate expressions. These kinetic models, translated to deposition rates for the three solid products Si, WSi$_2$, and W$_5$Si$_3$, are given as

$$
\begin{aligned}
G_1 &= k_{11}e^{(-k_{12}/T)}P_1P_2^2 \\
G_2 &= k_{21}e^{(-k_{22}/T)}\frac{P_1P_2}{1+k_{23}P_1} \\
G_3 &= k_{31}e^{(-k_{32}/T)}P_1P_2^{1/2}
\end{aligned}
\tag{1}
$$

where $G_i(.)$, $i = 1, 2, 3$ are the deposition (growth) rates of Si, WSi$_2$, W$_5$Si$_3$, respectively, P_1, P_2 are the partial pressures of the reactants WF$_6$, DCS and k_{ij} are constants, e.g., see [1, 2].

The typical objective in such a process is to adjust the manipulated variables (control inputs) $u = [T, P, F_1]^\mathsf{T}$ where T is the susceptor temperature, P is the total pressure and F_1 is the WF$_6$ flowrate, so as to maintain desired values for the total deposition rate and silicon-to-tungsten ratio, while minimizing the variation of the deposition rate across the wafer surface. Estimates of these quantities can be obtained given measurements at different points on the wafer surface. Here, assuming a radially symmetric reactor, we define the spatial nonuniformity (Nun) across the wafer surface as the standard deviation of growth rates at the different measurement points along the wafer radius, divided by average growth rate:

$$
\begin{aligned}
\text{Nun} &= \left[\frac{1}{N}\sum_{n=1}^{N}(GR_n - \overline{GR})^2\right]^{1/2} /\overline{GR} \\
GR_n &= \sum_{i=1}^{3}G_i(n) \\
\overline{GR} &= \frac{1}{N}\sum_{n=1}^{N}GR_n
\end{aligned}
\tag{2}
$$

where $G_i(n)$, GR_n are the deposition rate of the solid species i and the total deposition rate at the measurement point n, and $\overline{GR}$ is the spatially averaged deposition rate. Further, from the stoichiometry of the apparent reaction model, the silicon to tungsten ratio (Si/W) of the deposited film at the wafer center is given by

$$\text{Si/W} = \frac{A_1 G_1 + 2A_2 G_2 + 3A_3 G_3}{A_2 G_2 + 5A_3 G_3} \qquad (3)$$

where A_i denotes the density to molecular weight ratio of the solid product i.

The main difficulty in obtaining a simplified, control-oriented model of this process (relating the manipulated variables with the outputs of interest) is associated with the dependence of the local reaction and deposition rates on the unmeasured partial pressures of WF_6 and DCS on the wafer surface. To develop such a relationship, we observe that in a reaction-free environment the partial pressures P_1 and P_2 could be determined by the corresponding mole fractions and the total pressure; on the other hand, when a chemical reaction occurs, P_1 and P_2 at the wafer surface will drop due to the consumption of WF_6 and DCS during the reaction. Assuming that the amount of reduction of the partial pressure is proportional to the apparent overall reaction rate, for that species, we employ multiple response surface techniques to approximate these rates by an empirical expression of the manipulated variables. Thus, we arrive at the following empirical (but physically motivated) model, relating the manipulated variables with the reactant partial pressures at a point on the wafer surface:

$$
\begin{aligned}
P_1 &= \tfrac{\theta_{11}F_1 P}{F_1+A} - (\theta_{12} + \theta_{13}T + \theta_{14}P \\
&\quad + \theta_{15}TP + \theta_{16}TPF_1) \\
P_2 &= \tfrac{\theta_{21}AP}{F_1+A} - (\theta_{22} + \theta_{23}T + \theta_{24}P \\
&\quad + \theta_{25}TP + \theta_{26}TPF_1)
\end{aligned} \qquad (4)
$$

where θ_{jk} are model parameters and A is a constant equal to the fixed flow rate of DCS. Notice that the "reaction feedback" modeling idea is invoked in selecting the parametric structure of the above model. That is, the first term in equations (4) describes the reaction-free dependence of the partial pressures on the manipulated variables while the rest describe the effect of the partial pressure reduction due to the reactions.

Next, substituting the reactant partial pressure expressions (4) in the deposition rate, nonuniformity and ratio expressions (1), (2) and (3), we obtain a model of the quantities of interest in terms of the manipulated variables $u = [T, P, F_1]^\mathsf{T}$ and the vector of adjustable parameters $\theta = [\theta_{11}, \ldots, \theta_{16}, \theta_{21}, \ldots, \theta_{26}]^\mathsf{T}$. Notice that since the deposition models of all three solid products share the same set of parameters at each measurement point, the total number of parameters is significantly reduced compared to an approach that would model the deposition rate of the solid products individually.

Given the simplified model (2) and (3), with the various rate expressions determined by (1) and (4) the control problem may now be formulated as

$$\min_u \text{Nun}(u, \theta^*), \text{ s.t. } \begin{cases} \overline{GR}(u, \theta^*) - GR^* = 0 \\ \text{Si/W}(u, \theta^*) - \text{Si/W}^* = 0 \\ u_{min} \le u \le u_{max} \end{cases}$$

(5)

where GR^* and Si/W^* are desired values of the average deposition rate and Si/W ratio. Here θ^* is used to denote the model parameters for which the rate expressions (1) offer a "good" approximation of the actual process. Based on preliminary simulation studies, such a value exists but may (and typically does) depend on the processing conditions. Consequently, θ^* is only partially known, a situation that is expected to be even more pronounced in an actual processing system. In order to obtain an approximate solution to the above control problem, we adopt an indirect adaptive control approach whereby the estimates of the model parameters θ are updated using deposition rate measurements, so that $G_i(u, \theta; n) \simeq G_i(u, \theta^*; n)$ where $G_i(u, \theta^*; n)$ is the deposition rate of product species i at a point n (e.g., see [6, 7, 8]). In turn, these estimates are used in (5) instead of θ^* and the constrained minimization problem is solved to determine the control input for the next run.

The behavior of the above adaptive controller was tested by using the CFDSWR platform to simulate the actual process. The simulation results (available upon request) show that the trajectories of the average total deposition rate and Si/W ratio are controlled to their individual desired values (2000 A/min and 2, respectively) while the nonuniformity is "minimized."[1]

References

[1] G.B. Raupp, T.S. Cale, M.K. Jain, B. Rogers and D. Srinivas, "Step Coverage of Tungsten Silicide Films Deposited by Low Pressure Dichlorocilane Reduction of Tungsten Hexafluoride," *Thin Solid Films*, 193/194, pp. 234–243, 1990.

[2] Damodaran Srinivas, "Kinetics of LPCVD of Tungsten Silicide", M.S. Thesis, A.S.U., Dec. 1989.

[3] J.H. Park, *User's guide for CFDSWR: Computational Fluid Dynamics of a Single Wafer Reactor*. A.S.U. 1993.

[4] J.H. Park, *Simulation of LPCVD Using Combined Reactor-Scale and Feature-Scale Models*, Ph.D. thesis, A.S.U. Aug.1992.

[5] R.S. Guo, and E. Sachs, "Modeling, Optimization and Control of Spatial Uniformity in Manufacturing Processes," *IEEE Trans. Semicond. Manuf.*, 6, 1, 1993.

[6] S. Sastry, and M. Bodson, *Adaptive Control Stability, Convergence and Robustness*. Prentice Hall, 1989.

[7] K.S. Tsakalis, and P.E. Crouch, "A Simple Adaptive Controller for an Oxidation Process," *Proc. 1992 ACC*.

[8] K.S. Tsakalis and L. Song, "Set-membership estimation for weakly nonlinear models: An application to the adaptive control of semiconductor manufacturing process", *33rd CDC*, 1994.

---

[1] Note that the adaptive control algorithm minimizes the estimated model nonuniformity instead of the actual one; nevertheless, the resulting value of Nun is fairly close to the optimal.

Increasing Throughput in Low Pressure Chemical Vapor Deposition: An Optimal Control Approach*

Timothy S. Cale[†], Peter E. Crouch[‡], Lijuan Song[‡], and Kostas S. Tsakalis[‡]

[†]Center for Solid State Electronics Research   [‡] Center for System Science and Engineering

Arizona State University

Box 877606, Tempe, AZ 85287-7606

e-mail: Peter.Crouch@asu.edu

Abstract

The application of optimal control theory to the process of low pressure chemical vapor deposition on patterned surfaces can substantially decrease the processing time for a given step coverage, compared with the programmed rate chemical vapor deposition (PRCVD) process. The control model is developed from the simultaneous one-dimensional Knudsen diffusion and chemical reaction description. For such a model, the optimal control problem is formulated as to find a temperature trajectory yielding the minimum processing time and its solution is computed numerically via a modified variation of extremals method. For the thermally activated deposition of silicon dioxide from tetraethylorthosilicate (TEOS) and for a ninety-six percent step coverage, the optimal control-generated temperature trajectory results in time-savings of approximately twenty-eight percent, when compared to the PRCVD approach.

1. Introduction

Higher scales of device integration and cost reduction of integrated circuits are two major trends in the microelectronics industry. In order to achieve cost reduction, a common practice is to increase the device throughput. In turn, such an increase can be achieved by increasing the number of devices per wafer. Therefore, four inch wafer fabrication lines are being converted to five or six and even eight inch wafer lines. This conversion, however, can adversely affect the inter-wafer and intra-feature film thickness uniformity in low pressure chemical vapor deposition (LPCVD) processes. Consider, for example conventional axial-flow, volume-loaded, multiple wafer-in-tube reactors (MWRs) commonly used in LPCVD processes [2]. To promote inter-wafer uniformity, for any wafer size, source gas conversion must be maintained at a rather low level in order to minimize axial reactant concentration gradients. Low gas conversion leads to both economic inefficiency and potentially hazardous operations, since the source gases are often expensive and extremely toxic. To maintain intra-wafer uniformity in MWRs with the wafers perpendicular to the reactant flow, the deposition rate must be much lower than the characteristic transport rate from the wafer edge to center. Lower deposition rates result in longer processing time to deposit films of given thickness.

In an effort to avoid the uniformity problems in MWRs, the microelectronics industry has used single wafer reactors (SWRs). These have the advantage that much higher conversion of the expensive and toxic reactants can be obtained without violating uniformity constraints. These high deposition rates are necessary for SWRs to achieve throughput parity with MWRs with the same film product specifications or process constraints. An important issue arising in this approach is that the step coverage in patterned regions of wafer degrades as deposition rate is increased. Thus, maintaining high step coverage under high deposition rate is a big challenge in designing rapid thermal LPCVD processes. Cale and coworkers have suggested a novel programmed rate process protocol for a SWR LPCVD process (PRCVD) that can be used to decrease the required deposition time and hence increase throughput subject to a given step coverage constraint [1, 2].

In this paper, we employ optimal control theory to develop an alternative process protocol in an effort to minimize the processing time, subject to the same step coverage constraint. In a similar fashion as in the PRCVD approach, we use the processing temperature as the control, or manipulated, variable while keeping all other conditions constant. Our simulation results show that the temperature profile obtained by solving the associated optimal control problem can yield even further reduction of the processing time than the PRCVD approach.

2. Modeling

The continuum diffusion reaction model has been used in the recent literature to predict step coverage in LPCVD [1, 2]. Raupp and Cale [4] derived the equations describing time-dependent simultaneous heterogeneous reaction and Knudsen diffusion which apply to deposition through a single heterogeneous reaction of arbitrary kinetics in a feature of arbitrary symmetric cross section. The coordinate system for the feature is a moving coordinate system with the origin at the center of the feature mouth during deposition. Such a coordinate system for a trench is shown in Figure 1, where $X(t)$ is the film thickness in the bottom of the trench at time t; $L(t)$ is the film thickness at the mouth of the trench at time t; $Z_b = H(t)$ is the instantaneous trench depth; $W(Z, t)$ is the instantaneous width of the

Figure 1: Cross section of a feature during deposition

Dimensionless Variable	Definition
axial distance	$\xi = Z/H(t)$
time	$\tau = t D_{Ro}/H_0^2$
feature depth	$\mathcal{H}(\tau) = H(t)/H(0)$
feature cross sectional area	$\mathcal{A}(\xi,\tau) = \frac{A(Z,t)}{A(0,0)}$
feature perimeter	$\mathcal{P}(\xi,\tau) = \frac{P(Z,t)}{P(0,0)}$
feature width	$\mathcal{W}(\xi,\tau) = \frac{W(Z,t)}{W(0,0)}$
Knudsen diffusivity	$\mathcal{D}(\xi,\tau) = \frac{D_i(Z,t)}{D_i(0,0)}$
concentration	$\theta_i(\xi,\tau) = \frac{C_i(Z,t)}{C_i(0,0)}$
rate of reaction	$G_j(\xi,\tau) = \frac{R_j(P(Z,t),T(t))}{R_j(P(0,0),T(0))}$
solid density	$\rho_{S_k} = C_{S_k}/C_{Ro}$

Table 1: Definition of Dimensionless Variables
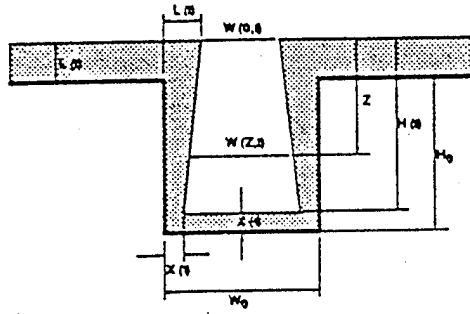
trench at depth Z; W_0 is the initial width of a trench; and H_0 is the initial trench depth.

We can assume [2] that deposition occurs under conditions such that the rate depends only on the concentration of the limiting reactant, the feature is spatially isothermal, and that surface diffusion and radial or lateral gas concentration gradients are negligible. It is, therefore, reasonable to treat the molecular transport as an one dimensional process in which species flux is expressed in terms of local concentration gradients and Knudsen diffusion coefficients. The expressions for Knudsen diffusivity are based on a cross-sectional average for idealized feature geometry, e.g., infinitely long rectangular trenches and cylindrical contact holes. Here, we only consider the infinitely long rectangular trench model, although the same methodology is applicable to cylindrical contact holes as well. For infinitely long rectangular trenches, [2, 6, 7] give the following estimate of D_i (instantaneous, cross sectional-averaged, local Knudsen diffusivity of a gaseous species i):

$$D_i(Z,t) = \left[\frac{8K_B T(t)}{\pi m_i}\right]^{0.5} \frac{H(t)}{4}\left[\frac{18 + 7\alpha(Z,t)}{18 + 16\alpha(Z,t) + 2\alpha^2(Z,t)}\right] \quad (1)$$

where K_B is the Boltzmann constant, T is the temperature (in K), m_i is the molecular mass of i, and α is the instantaneous, local aspect ratio, given by $\alpha(Z,t) = \frac{H(t)}{W(Z,t)}$. Using the notation:

• C_i: instantaneous, local concentration of species i;
• A and P: feature sectional area available for molecular flow and feature perimeter at position Z and time t, respectively;
• $p_i(Z,t)$: instantaneous partial pressure of species i at position Z and time t (in torr);
• v_{ji}: instantaneous rate of j-th heterogeneous chemical reaction based on local conditions;
• $C_i(0,t) = \frac{p_i(0,t)}{R_g T(t)}$ (R_g is the ideal gas constant);
• $D_i(Z_b,t)\frac{\partial C_i(Z_b,t)}{\partial Z} = \sum_{j=1}^{r} v_{ji} R_j(p_i(Z_b,t),T(t))$;
and the dimensionless variables defined in Table 1, the model equation in dimensionless from is described as follows [1]:

Balance of species i:

$$\frac{\partial[\mathcal{A}(\xi,\tau)\theta_i(\xi,\tau)]}{\partial\tau} = \frac{D_{io}}{D_{ro}\mathcal{H}^2(\tau)}\frac{\partial}{\partial\xi}\left[\mathcal{A}(\xi,\tau)\mathcal{D}(\xi,\tau)\frac{\partial\theta_i(\xi,\tau)}{\partial\xi}\right] + \lambda_{Ri}\mathcal{P}(\xi,\tau)\sum_{j=1}^{r} v_{ji}G_j(\xi,\tau)\phi_j(\tau) \quad (2)$$

Initial conditions:

$$\theta_i(\xi,0) = 1 \quad ; \quad H(0) = 1 \quad (3)$$

Boundary conditions:

$$\begin{aligned}\theta_i(0,\tau) &= \frac{p_i(0,t)T(0)}{p_i(0,0)T(t)} = \frac{C_i(0,t)}{C_i(0,0)} \\ \frac{\partial\theta_i(\xi_b,\tau)}{\partial\xi} &= \frac{D_{Ro}C_{Ro}}{D_o C_{io}}\frac{A(0,0)}{H(0)P(0,0)}\frac{\mathcal{H}(\tau)}{\mathcal{D}(\xi_b,\tau)} \\ & \quad \sum_{j=1}^{r} v_{ji}G_j(\xi_b,\tau)\phi_j(\tau)\end{aligned} \quad (4)$$

where $0 \le \xi \le \xi_b$, $\tau > 0$, and $\xi_b = 1$. ϕ_j is referred to as the step coverage modulus for the reaction j and is given by

$$\phi_j(\tau) = \frac{H^2(0)P(0,0)R_j(p(0,t),T(t))}{A(0,0)D_{Ro}C_{Ro}} \quad (5)$$

Finally λ_{Ri} denotes the partial pressure ratio and is defined as $\lambda_{Ri} = \frac{p_{Ro}}{p_{io}} = \frac{C_{Ro}}{C_{io}}$.

The LPCVD process of deposition of silicon dioxide by TEOS pyrolysis involves a single gaseous reactant and deposition of a single solid species SiO_2. The reaction stoichiometry and rate expression for this process are [8]: TEOS $\longrightarrow SiO_2$ + products

$$R_N(p,T) = k_o \exp\left(-\frac{E_a}{R_g T}\right)p^N, \quad \left(\frac{mol}{cm^2 s}\right) \quad (6)$$

where, the temperature T and partial pressure p have units of Kelvin and torr, respectively. The activation energy E_a is 195kJ/mol and $k_o = 38.77$ while the order of reaction is $N = 0.5$. Finally, in dimensionless form, the model equations for LPCVD of SiO_2 from TEOS in long rectangular trenches become [2, 4]:

Species balance for TEOS:

$$\frac{\partial\theta}{\partial\tau} = \frac{1}{\mathcal{W}\mathcal{H}^2}\frac{\partial}{\partial\xi}\left(\mathcal{D}\mathcal{W}\frac{\partial\theta}{\partial\xi}\right) - \frac{\phi G}{\mathcal{W}}\left(1 - \frac{\theta}{\rho}\right) \quad (7)$$

Boundary conditions:

$$\theta(0,\tau) = \frac{p(0,t)T(0)}{T(t)p(0,0)} \quad ; \quad \frac{\partial\theta(\xi_b,\tau)}{\partial\xi} = -\frac{\mathcal{H}(\tau)G(\xi_b,\tau)\phi(\tau)}{2\alpha_o\mathcal{D}(\xi_b,\tau)} \quad (8)$$

where G is the dimensionless rate of reaction, ϕ is the step coverage modulus given by

$$\phi(\tau) = \frac{2H^2(0)R(p(0,t),T(t))}{C(0,0)D(0,0)W(0,0)} \quad (9)$$

and α_0 is initial aspect ratio, given by $\alpha_0 = H_0/W_0$.

3. Programmed Rate CVD (PRCVD)

In common commercial deposition processes, the wafer temperature and partial pressures of all reactants at the wafer surface are usually held constant. Therefore, the rate of deposition remains constant at the wafer surface during the entire process. Such processes are referred to as constant rate chemical deposition (CR-CVD) processes [2, 3]. Since the process parameters remain constant in CRCVD, the model parameters (step coverage modulus and reactant partial pressure ratios at the wafer surface) also remain constant. The objective of a PRCVD process protocol is to vary the deposition rate as the deposition progresses, by changing one or more of the process parameters, so as to decrease the deposition time, subject to a specific final step coverage constraint. As mentioned in the Introduction, such a protocol is of particular interest in SWRs, though in principle the same concept can be applied to other reactor configurations [2]. The PRCVD process path employed in [1] consists of two distinct legs: (1) A programmed rate leg during which the deposition rate at the wafer surface is adjusted so as to keep the instantaneous differential step coverage Γ (see below) constant. (2) A constant rate leg during which the process is continued at a constant deposition rate. For the programmed rate portion of the process, the deposition rate is adjusted by changing the temperature since this is the variable having the most dominant effect on the reaction rate. The point at which the process is changed from the programmed rate leg to the constant rate leg, is referred to as the switching point (SP). The switching point is defined to be the time at which a specified percentage of feature closure is obtained. Therefore, a PRCVD process path is determined by selecting Γ and SP. A judicious selection of Γ and SP can lead to a higher average rate of deposition during the process and, hence, to significant savings in processing time compared with the corresponding CRCVD process yielding the same step coverage.

To demonstrate the use of the PRCVD approach, we consider the case of SiO_2 deposition from TEOS decomposition in a rectangular trench. Since the reaction rate at the wafer surface varies during PRCVD, the step coverage modulus ϕ also varies. Thus, the problem of determining a rate path is equivalent to finding a path for the step coverage modulus ϕ. For this process, ϕ can be expressed as [1]

$$\phi(\tau) = \frac{2H_o^2 R_g T_o}{L_o D_{io} P_{io}} R_N(p, T) \qquad (10)$$

The above equation reveals that the step coverage modulus can be varied by changing either the partial pressure of TEOS at the wafer surface, or the wafer temperature, or both, during the programmed rate leg. We only discuss the case where the wafer temperature is varying while maintaining the TEOS partial pressure at the wafer surface constant.

Further, we can assume [2] that at any particular instant in time, the gas-phase concentration profiles in the feature are the steady-state profiles that would exist for the reaction conditions prevailing at that instant and that the molar density of the gas is negligible in comparison to that of the solid film, i.e., $\theta \ll \rho$. Using these assumptions, the species balance for TEOS (7),

and the corresponding boundary conditions become:

$$\frac{\partial}{\partial \xi} \mathcal{D}(\xi, \tau) \mathcal{W}(\xi, \tau) \frac{\partial \theta(\xi, \tau)}{\partial \xi} - \phi(\tau) G(\xi, \tau) \mathcal{H}^2(\tau) = 0 \qquad (11)$$

$$\theta(0, \tau) = \frac{T(0)}{T(t)} \quad ; \quad \frac{\partial \theta(\xi_b, \tau)}{\partial \tau} = -\frac{\mathcal{H}(\tau)}{2\alpha_o} \phi(\tau) \frac{G(\xi_b, \tau)}{\mathcal{D}(\xi_b, \tau)} \qquad (12)$$

The reactant concentration profile in the feature at any instant in time can be determined by solving the above differential equation and is given by

$$\theta(\xi, \tau) = \theta(0, \tau) - \phi(\tau) \mathcal{H}^2(\tau) \left[\mathcal{G}(\xi_b, \tau) + \frac{\mathcal{W}(\xi_b, \tau) G(\xi_b, \tau)}{2\alpha_o \mathcal{H}(\tau)} \right]$$
$$\int_0^\xi \frac{ds}{\mathcal{D}(s, \tau) \mathcal{W}(s, \tau)} - \int_0^\xi \frac{\mathcal{G}(s, \tau) ds}{\mathcal{D}(s, \tau) \mathcal{W}(s, \tau)} \right]$$
$$(13)$$

where $\mathcal{G}(\xi, \tau) = \int_0^\xi G(s, \tau) ds$

The instantaneous differential step coverage Γ is now defined by $\Gamma = \frac{dX(\tau)}{d\tau} / \frac{dL(\tau)}{d\tau}$. Since $\dot{L} = R(p(Z, T), T(t))$, $\dot{X} = R(p(0, T), T(t))$ and $2X(t)L + W(Z, t) = W_0$ (see Fig. 1), we obtain

$$\Gamma = \left[\frac{\theta(\xi_b, \tau)}{\theta(0, \tau)} \right]^N \qquad (14)$$

During the programmed rate leg of the PRCVD, Γ is held constant and the temperature is determined by solving (10), (13) and (14) using a first order approximation. During the constant rate leg (after SP), the temperature is held constant and equal to the temperature achieved at the end of the programmed rate leg. A reduction of processing time with the PRCVD approach versus the CRCVD one, depends on the selected values of the process path parameters Γ and SP. For a given step coverage, however, Γ and SP may be found by trial and error so as to minimize processing time [1, 2].

4. Optimal Control CVD (OCCVD)

The first step in applying optimal control theory to the SiO_2 deposition by TEOS decomposition, is to describe the process model by a set of ordinary differential equations. Then, the optimal control problem for the CVD process becomes one of determining the temperature trajectory (in this case the control input) subject to the final step coverage constraints such that the processing time is minimized.

From the definition of the instantaneous step coverage Γ, and since the final step coverage is evaluated as the ratio of the final film thickness at the base of the feature and the final thickness at the wafer surface, we obtain the following simplified model of the process:

$$\dot{L} = R(p(Z, T), T) \quad ; \quad \dot{X} = \Gamma(t) \dot{L} \qquad (15)$$

The first order perturbation solution of this pseudo-steady-state model provides a reasonable estimate of the actual concentration and deposition profiles for the range of ϕ values (reaction rates) which yield uniform deposition over the wafer.[1] The first order perturbation solution to (13) for the dimensionless concentration profile in a trench can be written as

$$\theta(1, \tau) = \theta(0, \tau) - \frac{\phi(\tau) \mathcal{H}(\tau)(\alpha + 1)}{\mathcal{D}(0, \tau) H_0 \alpha_0} \qquad (16)$$

---

[1] Simulations of the full model (2), (7) support this claim [2].

where $\alpha = \alpha(Z_b, t)$. Combining (10), (13) and (14), we obtain the following expression for Γ

$$\Gamma(t) = \left[1 - \frac{4}{T_o}\left(\frac{\pi m_i}{8K_o}\right)^{1/2}T^{1/2}(\alpha+1)\varphi(\alpha)\frac{R(p(0,t),T(t))}{C_o}\right]^N \quad (17)$$

$$\varphi(\alpha) = \frac{18 + 16\alpha(Z,t) + 2\alpha^2(Z,t)}{18 + 7\alpha(Z,t)}$$

The above equation reveals that Γ depends only on the variables T and α, since $R(p(0,t),T(t))$ depends only on the temperature $T(t)$ when the partial pressure $p(0,t)$ is constant. Thus, the optimal control can be cast as follows:

Optimal Control Problem: *For the system model (15) find an optimal control T, to minimize the processing time t_f subject to the constraints*

$$X(0) = L(0) = 0 \; ; \; \frac{X(t_f)}{L(t_f)} \geq SC \; ; \; L(t_f) = L_* \quad (18)$$

where Γ is given by equation (17), SC, L_* are the desired step coverage and final thickness.

To solve the above optimal control problem we may employ numerical techniques to determine the optimal control trajectories as the solution to a two-point boundary-value problem. One such general method is the so-called Variation of Extremals [9]. However, since in its standard form this method is not directly applicable to our case, a suitable modification needs to be derived. To achieve this, let us begin by employing optimal control theory to determine necessary conditions for a control trajectory to be optimal. In general, the problem is to find an admissible control U^* in the feasible set Ω that causes the system

$$\dot{X}(t) = a(X(t), U(t), t)$$

to follow an admissible trajectory $X^*(t)$ minimizing the performance index

$$J(U) = h(X(t_f), t_f) + \int_{t_0}^{t_f} g(X(t), U(t), t)dt$$

and satisfies boundary conditions $X(t_0) = X_0$ and $X(t_f) = X_f$, where t_0 is the specified initial time and t_f is the unknown final time. In our case (minimum time problem), $h \equiv 0$ and $g \equiv 1$, while the control U is the temperature T.[2] By defining the Hamiltonian function $H \triangleq g + P^T a$, the necessary conditions for optimality are

$$\dot{X}^*(t) = \frac{\partial H}{\partial P}(X^*(t), U^*(t), P^*(t), t)$$
$$\dot{P}^*(t) = -\frac{\partial H}{\partial X}(X^*(t), U^*(t), P^*(t), t) \quad (19)$$
$$H^*(t) = \min_{U(t)\in\Omega} H(X^*(t), U(t), P^*(t), t)$$

for all $t \in [t_0, t_f]$, with boundary conditions

$$X^*(t_0) = X_0 \; ; \; X^*(t_f) = X_f \; ; \; H^*(t_f) = 0 \quad (20)$$

Since $P(t_0)$ is unknown, we can guess a value $P^0(t_0)$ for the initial costate and use it to numerically integrate (19) from t_0 to t_f. Under the variation of extremals approach, the observed values of $P(t_f)$ are

[2]The selection of the temperature as the control variable, instead of —e.g.— partial pressure, is considered here since it is easier to manipulate in practice and leads to simpler expressions.

then used to systematically adjust the guessed values of $P(t_0)$. One technique for making systematic adjustments of the initial costate values is based on Newtons method for finding roots of nonlinear equations [9]. Thus, given n state equations and n costate equations, the update law for $P(t_0)$ is given by

$$P^{(i+1)}(t_0) = P^{(i)}(t_0) - \left\{[\mathcal{P}_P(t_f)]_i\right\}^{-1}[P(t_f)]_i \quad (21)$$

$$\dot{\mathcal{P}}_X = \left[\frac{\partial^2 H}{\partial P \partial X}(t)\right]_i \mathcal{P}_X + \left[\frac{\partial^2 H}{\partial P^2}\right]_i \mathcal{P}_P \quad (22)$$

$$\dot{\mathcal{P}}_P = \left[\frac{\partial^2 H}{\partial^2 X}(t)\right]_i \mathcal{P}_X + \left[\frac{\partial^2 H}{\partial X \partial P}\right]_i \mathcal{P}_P \quad (23)$$

$$\mathcal{P}_X(t_0) = 0 \; ; \; \mathcal{P}_P(t_0) = I \quad (24)$$

where $\mathcal{P}_X(P^{(i)}(t_0), t)$, $\mathcal{P}_P(P^{(i)}(t_0), t)$ are the matrices of partial derivatives of the components of $X(t)$ and $P(t)$ with respect to each of the components of $P(t_0)$, evaluated at $P^{(i)}(t_0)$. Notice that $\mathcal{P}_P$ and $\mathcal{P}_X$ are needed only at the terminal time t_f, while the above derivations assume continuity of partial derivatives. The notation $[\cdot]_i$ means that the enclosed terms are evaluated on the i-th trajectory.

Since the $P^*(t_f)$ is not specified in our case, we cannot directly apply the standard variation of extremals algorithm. In order to develop a suitable modification, we consider the observed final state $X(t_f)$ as the result of the choice of initial costates $P(t_0)$ and final time t_f. We then guess $P^{(0)}(t_0)$ and $t_f^{(0)}$, and update them so that the observed $X(t_f)$ converges to the desired value. Thus, we finally obtain the following equations for updating $P^{(i+1)}(t_0)$ and $t_f^{(i+1)}$.

$$\Delta X_f(i) = \mathcal{P}_X \Delta P(i) + \frac{\partial X}{\partial t_f}\Delta t_f(i)$$
$$0 = \frac{\partial H}{\partial X}\Delta X_f(i) + \frac{\partial H}{\partial P}\mathcal{P}_P \Delta P(i)$$
$$\begin{bmatrix} \Delta P(i) \\ \Delta t_f(i) \end{bmatrix} = \begin{bmatrix} \mathcal{P}_X & \frac{\partial X}{\partial t_f} \\ \frac{\partial H}{\partial P}\mathcal{P}_P & 0 \end{bmatrix}^{-1} \begin{bmatrix} \Delta X_f(i) \\ \frac{\partial H}{\partial X}\Delta X_f(i) \end{bmatrix} \quad (25)$$

where $\Delta X_f(i) = X_f - X^{(i)}(t_f)$, $\Delta P(i) = P^{(i+1)}(t_0) - P^{(i)}(t_0)$, $\Delta t_f(i) = t_f^{(i+1)} - t_f^{(i)}$, and all quantities are evaluated at $t_f^{(i)}$, $P^{(i)}(t_0)$. The iteration is terminated when $\|X_f - X(t_f^{(i)})\| \leq \gamma$ is satisfied.

It should be emphasized that the above procedure of computing the optimal trajectory does not account for the possibility that, during the iterations, the instantaneous step coverage Γ may become a complex number, especially near closure. Although a systematic way to prevent this situation is feasible (e.g., by introducing an additional constraint in the optimal control problem), for the sake of simplicity we adopted a different approach. That is, we use a similar strategy as the PRCVD process in that we only find the optimal control during the first leg so as to minimize the corresponding time, while for the second leg the processing time is fixed and the temperature is held constant.

5. Results and Discussion

To compare the results of CRCVD, PRCVD and OCCVD, we consider the model of SiO_2 deposition by TEOS decomposition, as described by (15) with $N =$

0.5. The desired step coverage considered here is $SC = 0.96$. For best results in the PRCVD case, [1] suggests the following choice of parameters: $\Gamma = 0.9775$ for the programmed rate leg and $SP = 0.899$. With these values the temperature under PRCVD decreases until the switching point SP and remains constant thereafter. The instantaneous step coverage Γ is held constant before and decreases after the switching point. The total process time achieved under PRCVD, for 98 percent closure, is 386 sec. This value reflects significant savings over the CRCVD approach which, for the same SC and closure, requires 729 sec (Fig. 2).

On the other hand, under the OCCVD approach, we first select the time of the second leg time as 100 seconds.[3] With the temperature held constant in this leg, equal to the final temperature of the first leg in the previous iteration, (15) are integrated backwards in time to obtain the final state $X(t_f)$ at the end of the first leg. This final state is then substituted in (25) to yield the next estimate for $P(t_0)$ and t_f. These values are, in turn, used to integrate the state equations (19) forward in time with the input U (temperature) computed accordingly. The procedure is repeated until convergence is achieved. The OCCVD approach results in a first leg time of 178 seconds and a total time, for 98 percent closure, of 278 seconds. Compared to the PRCVD approach, OCCVD yields a 28 percent reduction of processing time (Fig. 2).

The above results demonstrate that significant savings in processing time can be obtained, without compromising the step coverage constraint, by using optimal control theory to compute temperature trajectories for a CVD process. Another advantage of employing optimal control theory in such problems is that it helps to reduce, if not eliminate, the "judicious selection" of critical parameters (e.g., Γ and SP in the PRCVD approach). On the other hand, the price paid for these improvements is related to the increased computational complexity of the solution and the difficulty of implementing the optimal trajectories on the actual process. Without attempting to completely resolve these issues here, we note that the former is of lesser significance, especially in view of the computational power of modern computers. Furthermore, the trajectory implementation problem can be alleviated to a large extend by appropriately designing the local (inner-loop) temperature controllers. In the same vein it should be pointed out that, in contrast to the PRCVD approach, optimal control theory can easily account for the constraints imposed by the temperature dynamics in the solution. In such a formulation, the inner-loop temperature set-point would be used as the control variable in the optimal control problem, with an additional constraint (in the form of a simple dynamical model or just a temperature derivative constraint) imposed by the bandwidth of the local temperature closed-loop. Thus, although the exact relationship between the desired and the actual wafer temperatures is very complicated or even unknown, the proper use of the local feedback controller can ensure the successful implementation of the optimal temperature trajectory. This issue is anticipated
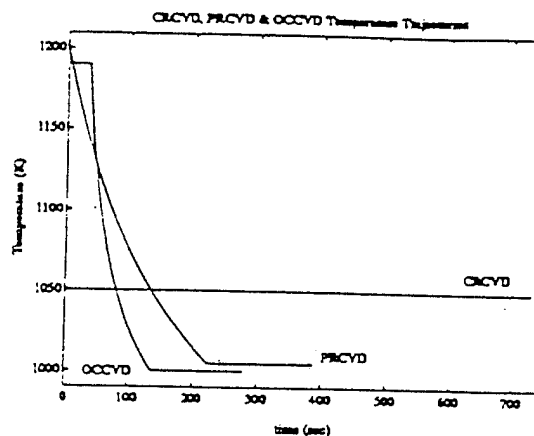

Figure 2: Temperature trajectory comparison under CRCVD, PRCVD and OCCVD

to be of importance in our future work, involving the experimental validation of the PRCVD and OCCVD approaches, since the maximum temperature rate capabilities of the available reactor (a "Spectrum 202") are below the unconstrained optimal trajectory rates.

Finally, a more subtle problem arises from the consideration of the "quality" of the deposited film in terms of grain size and adhesion properties (depending largely on the processing temperature). This issue is expected to impose constraints on the temperature range that need to be determined experimentally. In addition, in order to establish such an optimal trajectory generation procedure as a practical alternative to the current approaches, the various derivations and computations should be "automated" to a great extend in order to become usable by process engineers (e.g., provide the ability of easily generating a new optimal trajectory when the processing conditions change). These issues, however, extend beyond the scope of this study and are left as a topic of future research.

---
[3]This value is somewhat arbitrary and could be selected more systematically by performing an one-parameter optimization with respect to the second leg time. However, for the sake of simplicity, this approach is not adopted in the present study.

References

[1] M. K. Jain, *Maximization of Step Coverage at High Throughput During Low Pressure Deposition Processes.* Doctoral Thesis. Arizona State University, 1992.

[2] T. S. Cale, M. K. Jain and G. B. Raupp, "Programmed Rate Processing to Increase Throughput in LPCVD," *Journal of the Electrochemical Society*, Vol. 137, No. 5, may 1990.

[3] T. S. Cale, M. K. Jain and G. B. Raupp, "Tungsten and Other Advanced Metals for VLSI/ULSI Applications," *Materials Research Society*, Vol.V, Pittsburgh, PA, 1990. p.179.

[4] G. B. Raupp and T. S. Cale, *Chemistry of Materials*, 1, 207, 1989.

[5] J. H. Park, *Simulation of Low Pressure Chemical Vapor Deposition Using Combined Reactor Scale and Feature Scale Models.* Doctoral Thesis. Arizona State University, 1992.

[6] M. Knudsen, *The Kinetic Theory of Gases.* Wiley, New York, 1950.

[7] P. Clausing, *Physica.* 9, 65, 1929.

[8] A. C. Adams and C. D. Capio, *J. Electrochem. Soc.*, 126, 1042, 1979.

[9] D.E. Kirk, *Optimal Control Theory.* Prentice Hall, 1974.

OPTIMAL CONTROL FOR LPCVD

Timothy S. Cale†, Peter E. Crouch‡, Lijuan Song‡, and Kostas S. Tsakalis‡

†Center for Solid State Electronics Research
‡ Center for System Science and Engineering
Box 877606, Tempe, AZ 85287-7606
Arizona State University
Tempe, AZ 85287-6006

By employing an approximate model based on one-dimensional Knudsen diffusion and chemical reaction description of low-pressure chemical vapor deposition (LPCVD), an optimally controlled chemical vapor deposition (OCCVD) protocol is described, for the specific problem of maximizing throughput for a specified step coverage. The OCCVD approach is demonstrated by developing optimal temperature trajectories for the thermally activated deposition of silicon dioxide from tetraethoxysilane (TEOS). The results suggest that OCCVD is superior to programmed rate CVD (PRCVD) and increases the average deposition rate by more than a factor of two relative to constant rate CVD (CRCVD).

INTRODUCTION

Two major challenges the microelectronics industry faces today are the high scale of device integration and cost reduction of integrated circuits. One way to reduce cost is to increase device throughput. Step coverage and film thickness uniformity are critical constraints in LPCVD. Single-wafer reactors (SWRs) are becoming more common in order to obtain better uniformity. With the same film product specification or process constraints, high deposition rates are necessary for SWRs to maintain throughputs comparable to multiple-wafer reactor (MWRs). Unfortunately, the step coverage in patterned regions of wafers generally degrades as deposition rate is increased. Thus, the specification of process conditions which lead to acceptable step coverage and high deposition rate is a major goal of engineers designing LPCVD processes in SWRs. This problem is difficult to solve using conventional constant rate CVD (CRCVD), in which deposition conditions are held constant during the majority of the deposition process. After an initial transient, the deposition rate remains essentially constant at each location on the wafer surface during the process.

Symposium on Process Control, Diagnostics & Modeling in Semiconductor Manufacturing,

Reno, May 95

During deposition, feature aspect ratios increase, and the deposition rate during CR-CVD must be low enough to ensure good step coverage as features close, where the aspect ratios are the highest. Cale and coworkers have suggested programmed rate CVD (PRCVD) for SWR LPCVD processes, in which the operating conditions are changed during the deposition in a prescribed manner [1-3]. The deposition conditions are changed in PRCVD such that the deposition rate decreases during processing, as the feature aspect ratios increase. PRCVD can be used to decrease the deposition time for a given final step coverage, thereby increasing throughput.

In this paper, we employ optimal control theory to develop a new approach, referred to as optimally controlled chemical vapor deposition (OCCVD) for the specific problem of maximizing throughput for a specified step coverage. For the purpose of developing such a control strategy, we employ an approximate model based on a one-dimensional Knudsen diffusion and chemical reaction description [1-3]. The time savings achieved by OCCVD depend upon the chemistry and other process parameters, including the step coverage constraint. In this work, the temperature is changed during deposition. More generally, reactant flow rates and other process setpoints could be changed. Our simulation results show that the temperature trajectory obtained by solving the associated optimal control problem can yield even further reduction of the processing time than the PRCVD approach. Step coverages obtained by the OCCVD process were computed using the simulation package EVOLVE [4] in order to ensure that errors introduced through the use of the approximate model are relatively small.

MODELING

For the purpose of developing control strategies, we treat transport as a one dimensional process in which species fluxes are expressed in terms of local concentration gradients and Knudsen diffusion coefficients. Cale and coworkers [1-3] presented the equations which are appropriate for this simplistic model of transport and reaction at high Knudsen number. In this work, we assume that the deposition rate depends on a single species concentration, the features are spatially isothermal and there is no surface diffusion. We deal with deposition in trenches; however, the same methodology applies to vias as well. The origin of the coordinate system is at the center of the trench mouth, and moves with the surface during the deposition. Figure 1 shows such a trench, and

- $X(t)$ is the film thickness in the bottom of the trench at time t;

- $L(t)$ is the film thickness at the mouth of the trench at time t;

- $Z_b = H(t)$ is the instantaneous trench depth;

- $W(Z, t)$ is the instantaneous width of the trench at depth Z;

- W_0 is the initial width of a trench;

- H_0 is the initial trench depth.

The concentration of any species i at a depth Z in the feature is governed by

$$\frac{\partial}{\partial t}[A(Z,t)C_i(Z,t)] = \frac{\partial}{\partial Z}\left[A(Z,t)D_i(Z,t)\frac{\partial C_i(Z,t)}{\partial Z}\right] + P(Z,t)\sum_{j=1}^{r} v_{ji}R_j(p_i(Z,t),T(t))$$

$$(1)$$

where C_i is the instantaneous, local concentration of species i. A and P denote the area available for molecular flow and the feature perimeter at position Z and time t, respectively. $p_i(Z,t)$ is the instantaneous partial pressure of species i at position Z and time t while $T(t)$ denotes the temperature at the wafer surface at time t. R_j is the instantaneous rate of the j-th heterogeneous chemical reaction based on local conditions and v_{ji} is the generalized stoichiometric coefficient of species i in reaction j. D_i is the instantaneous, cross sectional-averaged, local Knudsen diffusivity of a gaseous species i. For infinitely long rectangular trenches, Ref. [3] gives the following estimate of D_i:

$$D_i(Z,t) = \left[\frac{8K_BT(t)}{\pi m_i}\right]^{0.5}\frac{H(t)}{4}\left[\frac{18+7\alpha(Z,t)}{18+16\alpha(Z,t)+2\alpha^2(Z,t)}\right] \qquad (2)$$

where K_B is Boltzmann's constant, T is the absolute temperature, m_i is the molecular mass of i, and α is the instantaneous, local aspect ratio given by

$$\alpha(Z,t) = \frac{H(t)}{W(Z,t)} \qquad (3)$$

The boundary conditions for the above second order partial differential equation are

$$C_i(0,t) = \frac{p_i(0,t)}{R_g T(t)} \qquad (4)$$
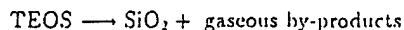
$$D_i(Z_b,t)\frac{\partial C_i(Z_b,t)}{\partial Z} = \sum_{j=1}^{r} v_{ji}R_j(p_i(Z_b,t),T(t)) \qquad (5)$$

where R_g is the ideal gas constant.

The LPCVD process of deposition of silicon dioxide by TEOS pyrolysis involves a single gaseous reactant and deposition of a single solid species SiO_2. The reaction stoichiometry and rate expression we use for this process are [3, 5]:

Dimensionless Variable	Definition
axial distance	$\xi = Z/H(t)$
time	$\tau = t D_{R0}/H_0^2$
feature depth	$\mathcal{H}(\tau) = H(t)/H(0)$
feature cross sectional area	$\mathcal{A}(\xi, \tau) = A(Z, t)/A(0, 0)$
feature perimeter	$\mathcal{P}(\xi, \tau) = P(Z, t)/P(0, 0)$
feature width	$\mathcal{W}(\xi, \tau) = W(Z, t)/W(0, 0)$
Knudsen diffusivity	$\mathcal{D}(\xi, \tau) = D_i(Z, t)/D_i(0, 0)$
concentration	$\theta_i(\xi, \tau) = C_i(Z, t)/C_i(0, 0)$
rate of reaction	$G_j(\xi, \tau) = R_j(P(Z, t), T(t))/R_j(P(0, 0), T(0))$
solid density	$\rho_{S_t} = C_{S_t}/C_{R0}$

Table 1: Definition of Dimensionless Variables

$$TEOS \longrightarrow SiO_2 + \text{gaseous by-products}$$

$$R_{(0.5)}(p, T) = 38.77 \exp\left(-\frac{195}{R_g T}\right) p^{0.5}, \quad \left(\frac{mol}{cm^2 s}\right) \tag{6}$$

where, the temperature T and partial pressure p have units of Kelvin and torr, respectively. The activation energy is in kJ/mol.

For the purposes of identifying the important parameters that dictate the step coverage and to determine their dependence on the CVD chemistry and operating conditions, the model equations are nondimensionalized [1-3, 5], using the definitions in Table 1. With these definitions, the model equations for LPCVD of SiO_2 from TEOS in long rectangular trenches become [3]:

Species balance for TEOS:

$$\frac{\partial \theta}{\partial \tau} = \frac{1}{\mathcal{W}\mathcal{H}^2}\frac{\partial}{\partial \xi}\left(\mathcal{D}\mathcal{W}\frac{\partial \theta}{\partial \xi}\right) - \frac{\phi G}{\mathcal{W}}\left(1 - \frac{\theta}{\rho}\right) \tag{7}$$

Boundary conditions:

$$\theta(0, \tau) = \frac{p(0, t)T(0)}{T(t)p(0, 0)} \tag{8}$$

$$\frac{\partial \theta(\xi_b, \tau)}{\partial \xi} = -\frac{\mathcal{H}(\tau)}{2\alpha_o}\phi(\tau)\frac{G(\xi_b, \tau)}{\mathcal{D}(\xi_b, \tau)} \tag{9}$$

where G is the dimensionless rate of reaction, ϕ is the step coverage modulus given by

$$\phi(\tau) = \frac{2H^2(0)R(p(0, t), T(t))}{C(0, 0)D(0, 0)W(0, 0)} \tag{10}$$

and α_0 is initial aspect ratio, given by $\alpha_0 = H_0/W_0$, and subscript b refers to the base of the feature.

We assume that at any particular instant in time, the gas-phase concentration profiles in the feature are the steady-state profiles that would exist for the reaction conditions prevailing at that instant and that the molar density of the gas is negligible in comparison to that of the solid film, i.e., $\theta \ll \rho$ [3]. Using these assumptions, the species balance for TEOS, given by equation (7), simplifies to:

$$\frac{\partial}{\partial \xi} \mathcal{D}(\xi,\tau) \mathcal{W}(\xi,\tau) \frac{\partial \theta(\xi,\tau)}{\partial \xi} - \phi(\tau) G(\xi,\tau) \mathcal{H}^2(\tau) = 0 \tag{11}$$

We discuss the simple case where the wafer temperature is varied while maintaining the TEOS partial pressure at the wafer surface constant. The boundary conditions for the above second order differential equation are

$$\theta(0,\tau) = \frac{T(0)}{T(t)} \tag{12}$$

and

$$\frac{\partial \theta(\xi_b,\tau)}{\partial \tau} = -\frac{\mathcal{H}(\tau)}{2\alpha_0} \phi(\tau) \frac{G(\xi_b,\tau)}{\mathcal{D}(\xi_b,\tau)} \tag{13}$$

The reactant concentration profile in the feature at any instant in time can be determined by solving the above differential equation and is given by

$$\theta(\xi,\tau) = \theta(0,\tau) -$$

$$\phi(\tau)\mathcal{H}^2(\tau)\left[\mathcal{G}(\xi_b,\tau) + \frac{\mathcal{W}(\xi_b,\tau)G(\xi_b,\tau)}{2\alpha_0\mathcal{H}(\tau)} \int_0^\xi \frac{d\xi}{\mathcal{D}(\xi,\tau)\mathcal{W}(\xi,\tau)} - \int_0^\xi \frac{\mathcal{G}(\xi,\tau)d\xi}{\mathcal{D}(\xi,\tau)\mathcal{W}(\xi,\tau)} \right] \tag{14}$$

where

$$\mathcal{G}(\xi,\tau) = \int_0^\xi G(\xi,\tau)d\xi \tag{15}$$

PROGRAMMED RATE CHEMICAL VAPOR DEPOSITION

The objective of a PRCVD process protocol is to vary the deposition rate as the deposition progresses, by changing one or more of the process parameters, so as to decrease the deposition time, subject to a specific final step coverage constraint. The PRCVD process path employed in References [3] and [5] consists of two distinct legs:

1. A programmed rate leg during which the deposition rate at the wafer surface is adjusted so as to keep the instantaneous differential step coverage Γ (see below) constant.

2. A constant rate leg during which the process is continued at a constant deposition rate.

The point at which the process is changed from the programmed rate leg to the constant rate leg, is referred to as the switching point (SP). The switching point is defined to be the time at which a specified percentage of feature closure is obtained. Therefore, a PRCVD process path is determined by selecting Γ and SP. A judicious selection of Γ and SP can lead to a higher average rate of deposition during the process and, hence, to significant savings in processing time compared with the corresponding CRCVD process yielding the same step coverage.

We demonstrate the use of this PRCVD approach by considering the case of SiO_2 deposition from TEOS decomposition in a rectangular trench. Since the reaction rate at the wafer surface varies during PRCVD, the step coverage modulus ϕ —which is proportional to the instantaneous deposition rate at the wafer surface, see equation (10) — also varies. Thus, the problem of determining a rate path is equivalent to finding a path for the step coverage modulus ϕ. For this process, ϕ can be expressed as

$$\phi(\tau) = \frac{2H_0^2 R_g T_0}{L_0 D_{io} P_{io}} k_0 \exp\left(-\frac{E_a}{R_g T}\right) p^{0.5}, \tag{16}$$

The above equation reveals that the step coverage modulus can be varied by changing either the partial pressure of TEOS at the wafer surface, or the wafer temperature, or both, during the programmed rate leg. In this work, the deposition rate is adjusted by changing the temperature, since this has a large effect on reaction rate.

The instantaneous differential step coverage Γ is defined by $\Gamma = \frac{dX(\tau)}{d\tau} / \frac{dL(\tau)}{d\tau}$. Because $\dot{L} = R(p(Z,T),T(t))$, $\dot{X} = R(p(0,T),T(t))$ and $2X(t) + W(Z,t) = W_0$, see Figure 1, we obtain

$$\Gamma = \left[\frac{\theta(\xi_b,\tau)}{\theta(0,\tau)}\right]^{0.5} \tag{17}$$

Since a path conforming to the PRCVD during the programmed rate leg of the process corresponds to constant Γ, equation (14) results in the following expression for $\phi(\tau)$.

$$\phi(\tau) = \frac{\theta(0,\tau)(1-\Gamma^2)}{\mathcal{H}^2(\tau)[\mathcal{G}(\xi_b,\tau) + \frac{W(\xi_b,\tau)G(\xi_b,\tau)}{2a_0\mathcal{H}(\tau)}\int_0^{\xi_b}\frac{d\xi}{\mathcal{D}(\xi,\tau)W(\xi,\tau)} - \int_0^{\xi_b}\frac{\mathcal{G}(\xi,\tau)d\xi}{\mathcal{D}(\xi,\tau)W(\xi,\tau)}]} \tag{18}$$

Thus, during the programmed rate leg of PRCVD, the temperature is determined from equations (18) and (16), using a first order approximation described in Ref. 5. During the constant rate leg (after SP), the temperature is held constant and equal to the temperature achieved at the end of the programmed rate leg. A reduction of processing time using PRCVD versus CRCVD depends on the selected values of the process path parameters Γ and SP. For a given step coverage, however, Γ and SP may be found by trial and error so as to minimize processing time [3, 5]. One such method is to first fix Γ and then choose the value of SP which achieves the desired final step coverage. A minimum time is then obtained by searching over the range of Γ for the one yielding the minimum total time. Figure 2 shows one of the possible PRCVD paths for SiO$_2$ deposition by TEOS decomposition.

OPTIMALLY CONTROLLED CHEMICAL VAPOR DEPOSITION

Problem Formulation

To apply optimal control theory to the CVD process, we describe the process model as a set of ordinary differential equations. Since the step coverage is evaluated as the ratio of the thickness at the base of feature to the thickness at the wafer surface, we use the equations which describe the growth rate at the wafer surface and the growth rate at the base of the trench as the process model. For SiO$_2$ deposition by TEOS deposition, the simplified model of the process is written as

$$\frac{dL(t)}{dt} = R = 38.77 \exp\left(-\frac{E_a}{R_g T}\right) p^{0.5}(0, t) \tag{19}$$

$$\frac{dX(t)}{dt} = \Gamma(t)\frac{dL(t)}{dt} = \Gamma(t)38.77 \exp\left(-\frac{E_a}{R_g T}\right) p^{0.5}(0, t) \tag{20}$$

where equation (20) comes from the definition of the instantaneous step coverage Γ.

We can obtain the expression for Γ by using the first order perturbation solution. The first order perturbation solution of the pseudo-steady-state model provides a reasonable estimate of concentration and deposition profiles for the range of ϕ values (reaction rates) which yield conformal depositions, since the concentration does not vary by more than a few percent from feature mouth to feature base [3,5].

The first order perturbation solution to equation (14) for the dimensionless concentration profile in a trench can be written as

$$\theta(1, \tau) = \theta(0, \tau) - \frac{\phi \mathcal{H}(\tau)(\alpha + 1)}{\mathcal{D}(0, \tau) H_0 \alpha_0} \tag{21}$$

where $\alpha = \alpha(Z_b, t)$. Combining equations (14), (18) and (21), we obtain the following expression for Γ

$$\Gamma(t) = \left[1 - \frac{4}{T_0} \left(\frac{\pi m_i}{8 K_b} \right)^{\frac{1}{2}} T^{\frac{1}{2}} (\alpha + 1) \varphi(\alpha) \frac{R(p(0,t), T(t))}{C_0} \right]^N \tag{22}$$

where

$$\varphi(\alpha) = \frac{18 + 16\alpha(Z,t) + 2\alpha^2(Z,t)}{18 + 7\alpha(Z,t)} \tag{23}$$

The above equations reveal that Γ depends on the variables T and α, since $R(p(0,t), T(t))$ depends only on the temperature $T(t)$ when the partial pressure $p(0,t)$ is constant. Thus, the optimal control can be cast as follows:

For the system model equations (19) and (20), find an optimal control T (temperature), to minimize the processing time t_f subject to the constraints

$$X(0) = L(0) = 0 \tag{24}$$

$$\frac{X(t_f)}{L(t_f)} \geq \text{ a given step coverage} \tag{25}$$

$$L(t_f) = \text{ a given thickness} \tag{26}$$

where Γ is given by equation (22).

Method

Since the system model is described by nonlinear ordinary differential equations, we do not attempt to solve the nonlinear optimal control problem analytically. Instead, we employ numerical techniques to determine the optimal control trajectories as the solution to a two-point boundary-value problem. One such general method is the so-called variation of extremals [6]. However, since in its standard form this method is not directly applicable to solve the optimal control for the CVD process, a suitable modification needs to be derived. In order to achieve this, we begin by employing optimal control theory to determine necessary conditions for a control trajectory to be optimal. In general, the problem is to find an admissible control U^* in the feasible set Ω, that causes the system

$$\dot{X}(t) = a(X(t), U(t), t) \tag{27}$$

to follow an admissible trajectory $X^*(t)$ that minimizes the performance index

$$J(U) = h(X(t_f), t_f) + \int_{t_0}^{t_f} g(X(t), U(t), t) dt \tag{28}$$

and satisfies boundary conditions $X(t_0) = X_0$ and $X(t_f) = X_f$, where t_0 is the specified initial time and t_f is the unknown final time.

By defining the Hamiltonian function

$$\mathcal{H}(X(t), U(t), P(t), t) = g(X(t), U(t), t) + P^T(t)[a(X(t), U(t), t)] \qquad (29)$$

the necessary conditions for optimality can be written as follows [6]

$$\dot{X}^*(t) = \frac{\partial \mathcal{H}}{\partial P}(X^*(t), U^*(t), P^*(t), t) \qquad (30)$$

$$\dot{P}^*(t) = -\frac{\partial \mathcal{H}}{\partial X}(X^*(t), U^*(t), P^*(t), t) \qquad (31)$$

$$\mathcal{H}(X^*(t), U^*(t), P^*(t), t) = \min_{U(t)\in\Omega} \mathcal{H}(X^*(t), U(t), P^*(t), t) \qquad (32)$$

for all $t \in [t_0, t_f]$, with boundary conditions

$$X^*(t_0) = X_o \qquad (33)$$

$$X^*(t_f) = X_f \qquad (34)$$

$$\mathcal{H}(X^*(t_f), U^*(t_F), P^*(t_f), t_f) + \frac{\partial h}{\partial t}(X^*(t_f), t_f) = 0 \qquad (35)$$

In our problem, we are interested in finding an optimal temperature trajectory to minimize the processing time. Hence, the performance index is the processing time while the control is the temperature T, i.e.,

$$J(U) = \int_{t_0}^{t_f} dt \qquad (36)$$

where $U = T$. Comparing with equation (38), we obtain the following equations $h(X(t_f), t_f) = 0$ and $g(X(t), U(t), t) = 1$ and the boundary conditions

$$X^*(t_0) = X_0 \qquad (37)$$

$$X^*(t_f) = X_f \qquad (38)$$

$$\mathcal{H}(X^*(t_f), U^*(t_f), P^*(t_f), t_f) = 0 \qquad (39)$$

Since the $P^*(t_f)$ is not specified in our case, we cannot directly apply the previously mentioned variation of extremals algorithm. In order to develop a suitable modification, we consider the observed final state $X(t_f)$ as the result of the choice of initial costates $P(t_0)$ and final time t_f. We then guess $P^{(0)}(t_0)$ and $t_f^{(0)}$, and update them so that the observed $X(t_f)$ converges to the desired value. Thus, we finally obtain the following equations for updating $P^{(i+1)}(t_0)$ and $t_f^{(i+1)}$.

$$X_f - X^{(i)}(t_f) = \frac{\partial X}{\partial P_0}(t_f^{(i)}, P_0^{(i)})(P_0^{(i+1)} - P_0^{(i)}) + \frac{\partial X}{\partial t_f}(t_f^{(i)}, P_0^{(i)})(t_f^{(i+1)} - t_f^{(i)}) \quad (40)$$

$$0 = \frac{\partial \mathcal{H}}{\partial X}[X_f - X(t_f^{(i)})] + \frac{\partial \mathcal{H}}{\partial P}\frac{\partial P}{\partial P_0}(t_f^{(i)}, P_0)(P_0^{(i+1)} - P_0^{(i)}) \quad (41)$$

$$\begin{bmatrix} P_0^{(i+1)} - P_0^{(i)} \\ t_f^{(i+1)} - t_f^{(i)} \end{bmatrix} = \begin{bmatrix} \frac{\partial X}{\partial P_0}(t_f^{(i)}, P_0^{(i)}) & \frac{\partial X}{\partial t_f}(t_f^{(i)}, P_0^{(i)}) \\ \frac{\partial \mathcal{H}}{\partial P}\frac{\partial P}{\partial P_0}(t_f^{(i)}, P_0) & 0 \end{bmatrix}^{-1} \begin{bmatrix} X_f - X^{(i)}(t_f) \\ \frac{\partial \mathcal{H}}{\partial X}[X_f - X(t_f^{(i)})] \end{bmatrix} \quad (42)$$

The iteration is terminated when $\|X_f - X(t_f^{(i)})\| \leq \gamma$ is satisfied.

It should be emphasized that the above procedure for computing the optimal trajectory does not account for the possibility that, during the iterations, the instantaneous step coverage Γ may become a complex number, especially near closure. Although a systematic way to prevent this situation is feasible (e.g., by introducing an additional constraint in the optimal control problem), for the sake of simplicity we adopted a different approach. We use a strategy similar to that used for PRCVD in that we only find the optimal control during the first leg so as to minimize the corresponding time, while for the second leg the processing time is fixed and the temperature is held constant.

RESULTS AND DISCUSSION

To compare the results of CRCVD, PRCVD and OCCVD, let us consider the model of SiO$_2$ deposition by TEOS decomposition, as described by Equations (19-20). The desired step coverage (SC) considered here is $SC = 0.96$. For best results in the PRCVD case, Reference [5] suggests the following choice of parameters: $\Gamma = 0.9775$ during the programmed rate leg and $SP = 0.899$. With these values and as shown in Figure 2, the temperature decreases until the switching point SP and remains constant thereafter. The total deposition time achieved using PRCVD, for 98 percent closure, is 386 seconds. For the same SC and percent closure, CRCVD requires 729 seconds of deposition time. The PRCVD process achieves 47% savings in processing time, compared with the CRCVD process which provides the same step coverage.

For OCCVD, we first select the time of the second leg as 100 seconds. This value is somewhat arbitrary and could be selected more systematically by performing an one-parameter optimization with respect to the second leg time. However, for the sake of simplicity, this approach is not adopted in the present study. With the

temperature held constant in this leg, equal to the final temperature of the first leg in the previous iteration, equations (19–20) are integrated backwards in time to obtain the final state $X(t_f)$ at the end of the first leg. This final state is then substituted in equation (42) to yield the next estimate for $P(t_0)$ and t_f. These values are, in turn, used to integrate the state equations (30–31) forward in time with the input U_2 (temperature) computed from equation (32). The procedure is repeated until convergence is achieved. The results of the OCCVD approach are shown in Figure 3. The first leg time is 178 seconds and the total time, for 98 percent closure, is 278 seconds. Compared to the PRCVD approach, OCCVD yields a 28% reduction of deposition time. The deposition time savings is 62% when compared with CRCVD (see Figure 4).

These estimated process time savings are validated by simulating the OCCVD, PRCVD and CRCVD process paths using the EVOLVE software [4]. The simulation results (Figure 5) show that the step coverage predicted using the simplified model differs from the more accurate EVOLVE predictions by 2.2% for OCCVD, 6.8% for PRCVD, and 4.6% for CRCVD. These results indicate that the time savings predicted using the simplified model remains qualitatively the same when a more accurate process model is used. The above results demonstrate that significant savings in processing time can be obtained, by using optimal control theory to compute temperature trajectories for a CVD process.

In practice, the time varying temperature predicted by OCCVD, will also affect partial pressures. As described in Ref. [7], reactor scale simulation could be used to determine the interaction between temperatures and partial pressures, and to establish reactor set-point trajectories. In contrast to the PRCVD approach, optimal control theory can easily account for the constraints imposed by the temperature dynamics in the solution (for example, to introduce constrains on the derivative of the temperature). These issues are left as topics of future research.

ACKNOWLEDGMENTS

REFERENCES

1. T. S. Cale, M. K. Jain and G. B. Raupp, Thin Solid Films, 193, 51 (1990).

2. T. S. Cale, M. K. Jain and G. B. Raupp, in Tungsten and Other Advanced Metals for VLSI/ULSI Application, S. S. Wong and S. Furukawa, Editors. p. 371, Material Research Society (1990).

3. T. S. Cale, M. K. Jain and G. B. Raupp, J. Electrochem. Soc., 137, 5, (1990).

4. T. S. Cale, EVOLVE 4.0b, low pressure deposition and etch process simulator, developed with support from the Semiconductor Research Corporation, the National Science Foundation and Motorola(1994).

5. M. K. Jain, Maximization of Step Coverage at High Throughput During Low Pressure Deposition Processes. Doctoral Thesis. Arizona State University(1992).

6. D.E. Kirk, Optimal Control Theory, Prentice Hall(1974).

7. J. H. Park and T. S. Cale, in Proceeding of the First International Conference on Rapid Thermal Processing, p. 393 (1993).

Optimal Control Processing to Increase Single Wafer Reactor Throughput In LPCVD*

Peter E. Crouch[†], Lijuan Song[†], Kostas S. Tsakalis[†], and Timothy S. Cale[‡]

Arizona State University

[†] Center for System Science and Engineering

[‡] Center for Solid State Electronics Research

Box 877606, Tempe, AZ 85287-7606

c/o E-mail: Peter.Crouch@asu.edu

Abstract

In this paper, Optimal Control theory is applied to develop an alternative process protocol in single wafer reactor LPCVD on patterned wafer in an effort to minimize the processing time, for given final step coverage. To achieve this, the operating conditions are changed during the deposition in a prescribed manner. A simplified control model is developed from the simultaneous one-dimensional Knudsen diffusion and chemical reaction description. The optimal control problem is formulated to find a temperature trajectory yielding the minimum processing time and its solution is computed numerically via a modified variation of extremals method. To demonstrate the concept of optimal control CVD (OCCVD), we consider the thermally activated deposition of silicon dioxide (SiO$_2$) from tetraethylorthosilicate (TEOS). Using the simplified control model, the estimated process time to achieve a 96% step coverage at 98% closure with the constant rate CVD (CRCVD) strategy is 729 seconds. Under the same conditions, the optimal control CVD (OCCVD) process time is 278 seconds. Compared to CRCVD, the process time saved with OCCVD is 62%.

1 Introduction

With the increasing demand for larger wafer diameters, single wafer reactors are preferred over volume-loaded multiple-wafer reactors, since they offer better deposition uniformity. Step coverage and film thickness uniformity are critical constraints in low pressure chemical vapor deposition (LPCVD) of patterned wafers. In order to maintain high device throughputs in single wafer reactors, the deposition rate must be much higher compared to multiple wafer reactors. An important issue arising here is that in general the step coverage deteriorates with increasing deposition rate. Thus, maintaining high step coverage under high deposition rate is a big challenge in designing rapid thermal LPCVD processes. This problem is difficult to solve using conventional constant rate CVD (CRCVD), in which deposition conditions are held constant during the majority of the deposition process. After an initial transient, the deposition rate remains essentially constant at each location on the wafer surface during the process. During deposition, feature aspect ratios increase, and the deposition rate during CRCVD must be low enough to ensure good step coverage as features close, where the aspect ratios are the highest. Cale and coworkers have suggested programmed rate CVD (PRCVD) for SWR LPCVD processes, in which the operating conditions are changed during the deposition in a prescribed manner [1-3]. The deposition conditions are changed in PRCVD such that the deposition rate decreases during processing, as the feature aspect ratios increase. PRCVD can be used to decrease the deposition time for a given final step coverage, thereby increasing throughput.

In this paper, we employ optimal control theory to develop a new approach, referred to as optimally controlled chemical vapor deposition (OCCVD) for the specific problem of maximizing throughput for a specified step coverage. For the purpose of developing such a control strategy, we employ an approximate model based on a one-dimensional Knudsen diffusion and chemical reaction description [1-3]. The time savings achieved by OCCVD depend upon the chemistry and other process parameters, including the step coverage constraint. In this work, the temperature is changed during deposition. More generally, reactant flow rates and other process set-points could be

Figure 1: Cross section of a feature during deposition

Dimensionless Variable	Definition
axial distance	$\xi = Z/H(t)$
time	$\tau = tD_{RO}/H_0^2$
feature depth	$\mathcal{H}(\tau) = H(t)/H(0)$
feature cross sectional area	$\mathcal{A}(\xi, \tau) = \frac{A(Z,t)}{A(0,0)}$
feature perimeter	$\mathcal{P}(\xi, \tau) = \frac{P(Z,t)}{P(0,0)}$
feature width	$\mathcal{W}(\xi, \tau) = \frac{W(Z,t)}{W(0,0)}$
Knudsen diffusivity	$\mathcal{D}(\xi, \tau) = \frac{D_i(Z,t)}{D_i(0,0)}$
concentration	$\theta_i(\xi, \tau) = \frac{C_i(Z,t)}{C_i(0,0)}$
rate of reaction	$G_j(\xi, \tau) = \frac{R_j(P(Z,t),T(t))}{R_j(P(0,0),T(0))}$
solid density	$\rho_{S_k} = C_{S_k}/C_{RO}$

Table 1: Definition of Dimensionless Variables

changed. Our simulation results show that the temperature trajectory obtained by solving the associated optimal control problem can yield even further reduction of the processing time than the PRCVD approach. Step coverages obtained by the OCCVD process were computed using the simulation package EVOLVE [4] in order to ensure that errors introduced through the use of the approximate model are relatively small.

2 Modeling

For the purpose of developing control strategies, we treat transport as a one dimensional process in which species fluxes are expressed in terms of local concentration gradients and Knudsen diffusion coefficients. Cale and coworkers [1-3] presented the equations which are appropriate for this simplistic model of transport and reaction at high Knudsen number. In this work, we assume that the deposition rate depends on a single species concentration, the features are spatially isothermal and there is no surface diffusion. We deal with deposition in trenches; however, the same methodology applies to vias as well. The origin of the coordinate system is at the center of the trench mouth, and moves with the surface during the deposition. Figure 1 shows such a trench, and $X(t)$ is the film thickness in the bottom of the trench at time t; $L(t)$ is the film thickness at the mouth of the trench at time t; $Z_b = H(t)$ is the instantaneous trench depth; $W(Z, t)$ is the instantaneous width of the trench at depth Z; W_0 is the initial width of a trench; and H_0 is the initial trench depth.

The concentration of any species i at a depth Z in the feature is governed by

$$\frac{\partial}{\partial t}[A(Z,t)C_i(Z,t)] = \frac{\partial}{\partial Z}\left[A(Z,t)D_i(Z,t)\frac{\partial C_i(Z,t)}{\partial Z}\right] +$$

$$P(Z,t)\sum_{j=1}^{r} v_{ji}R_j(p_i(Z,t),T(t)) \quad (1)$$

where C_i is the instantaneous, local concentration of species i. A and P denote the area available for

molecular flow and the feature perimeter at position Z and time t, respectively. $p_i(Z, t)$ is the instantaneous partial pressure of species i at position Z and time t while $T(t)$ denotes the temperature at the wafer surface at time t. R_j is the instantaneous rate of the j-th heterogeneous chemical reaction based on local conditions and v_{ji} is the generalized stoichiometric coefficient of species i in reaction j. D_i is the instantaneous, cross sectional-averaged, local Knudsen diffusivity of a gaseous species i. For infinitely long rectangular trenches, Ref. [3] gives the following an estimate of D_i. The boundary conditions for the above second order partial differential equation are

$$C_i(0,t) = \frac{p_i(0,t)}{R_g T(t)} \quad (2)$$

$$D_i(Z_b,t)\frac{\partial C_i(Z_b,t)}{\partial Z} = \sum_{j=1}^{r} v_{ji}R_j(p_i(Z_b,t),T(t)) \quad (3)$$

where R_g is the ideal gas constant.

The LPCVD process of deposition of silicon dioxide by TEOS pyrolysis involves a single gaseous reactant and deposition of a single solid species SiO_2. The reaction stoichiometry and rate expression we use for this process are [3, 5]:

$$TEOS \longrightarrow SiO_2 + \text{gaseous by-products}$$
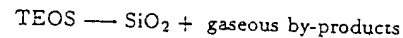
$$R_{(0.5)}(p,T) = 38.77 \exp\left(-\frac{195}{R_g T}\right) p^{0.5}, \quad \left(\frac{mol}{cm^2 s}\right) \quad (4)$$

where, the temperature T and partial pressure p have units of Kelvin and torr, respectively. The activation energy is in kJ/mol.

For the purposes of identifying the important parameters that dictate the step coverage and to determine their dependence on the CVD chemistry and operating conditions, the model equations are nondimensionalized [1-3, 5], using the definitions in Table 1. With these definitions, the model equations for LPCVD of SiO_2 from TEOS in long rectangular trenches become [3]:

Species balance for TEOS:

$$\frac{\partial \theta}{\partial \tau} = \frac{1}{\mathcal{W}\mathcal{H}^2}\frac{\partial}{\partial \xi}(\mathcal{D}\mathcal{W}\frac{\partial \theta}{\partial \xi}) - \frac{\phi G}{\mathcal{W}}(1-\frac{\theta}{\rho}) \qquad (5)$$

Boundary conditions:

$$\theta(0,\tau) = \frac{p(0,t)T(0)}{T(t)p(0,0)} \qquad (6)$$

$$\frac{\partial \theta(\xi_b,\tau)}{\partial \xi} = -\frac{\mathcal{H}(\tau)}{2\alpha_0}\phi(\tau)\frac{G(\xi_b,\tau)}{\mathcal{D}(\xi_b,\tau)} \qquad (7)$$

where G is the dimensionless rate of reaction, ϕ is the step coverage modulus given by

$$\phi(\tau) = \frac{2H^2(0)R(p(0,t),T(t))}{C(0,0)D(0,0)W(0,0)} \qquad (8)$$

and α_0 is initial aspect ratio, given by $\alpha_0 = H_0/W_0$, and subscript b refers to the base of the feature.

We assume that at any particular instant in time, the gas-phase concentration profiles in the feature are the steady-state profiles that would exist for the reaction conditions prevailing at that instant and that the molar density of the gas is negligible in comparison to that of the solid film, i.e., $\theta \ll \rho$ [3]. Using these assumptions, the species balance for TEOS, given by equation (7), simplifies to:

$$\frac{\partial}{\partial \xi}\mathcal{D}(\xi,\tau)\mathcal{W}(\xi,\tau)\frac{\partial \theta(\xi,\tau)}{\partial \xi} - \phi(\tau)G(\xi,\tau)\mathcal{H}^2(\tau) = 0 \qquad (9)$$

We discuss the simple case where the wafer temperature is varied while maintaining the TEOS partial pressure at the wafer surface constant. The boundary conditions for the above second order differential equation are

$$\theta(0,\tau) = \frac{T(0)}{T(t)} \qquad (10)$$

and

$$\frac{\partial \theta(\xi_b,\tau)}{\partial \tau} = -\frac{\mathcal{H}(\tau)}{2\alpha_0}\phi(\tau)\frac{G(\xi_b,\tau)}{\mathcal{D}(\xi_b,\tau)} \qquad (11)$$

The reactant concentration profile in the feature at any instant in time can be determined by solving the above differential equation and is given by

$$\theta(\xi,\tau) = \theta(0,\tau) - \phi(\tau)\mathcal{H}^2(\tau)\left[\mathcal{G}(\xi_b,\tau) + \frac{\mathcal{W}(\xi_b,\tau)G(\xi_b,\tau)}{2\alpha_0\mathcal{H}(\tau)}\right.$$
$$\left.\int_0^\xi \frac{ds}{\mathcal{D}(s,\tau)\mathcal{W}(s,\tau)} - \int_0^\xi \frac{\mathcal{G}(s,\tau)ds}{\mathcal{D}(s,\tau)\mathcal{W}(s,\tau)}\right]$$
$$\qquad (12)$$

where $\mathcal{G}(\xi,\tau) = \int_0^\xi G(s,\tau)ds$

3 Programmed Rate Chemical Vapor Deposition (PRCVD)

The objective of a PRCVD process protocol is to vary the deposition rate as the deposition progresses, by changing one or more of the process parameters, so as to decrease the deposition time, subject to a specific final step coverage constraint. The PRCVD process path employed in References [3] and [5] consists of two distinct legs: 1. A programmed rate leg during which the deposition rate at the wafer surface is adjusted so as to keep the instantaneous differential step coverage Γ (see below) constant. 2. A constant rate leg during which the process is continued at a constant deposition rate.

The point at which the process is changed from the programmed rate leg to the constant rate leg, is referred to as the switching point (SP). The switching point is defined to be the time at which a specified percentage of feature closure is obtained. Therefore, a PRCVD process path is determined by selecting Γ and SP. A judicious selection of Γ and SP can lead to a higher average rate of deposition during the process and, hence, to significant savings in processing time compared with the corresponding CRCVD process yielding the same step coverage.

We demonstrate the use of this PRCVD approach by considering the case of SiO_2 deposition from TEOS decomposition in a rectangular trench. Since the reaction rate at the wafer surface varies during PRCVD, the step coverage modulus ϕ —which is proportional to the instantaneous deposition rate at the wafer surface, see equation (8) — also varies. Thus, the problem of determining a rate path is equivalent to finding a path for the step coverage modulus ϕ. For this process, ϕ can be expressed as

$$\phi(\tau) = \frac{2H_0^2 R_g T_0}{L_0 D_{i0} P_{i0}}k_0 \exp\left(-\frac{E_a}{R_g T}\right)p^{0.5}, \qquad (13)$$

The above equation reveals that the step coverage modulus can be varied by changing either the partial pressure of TEOS at the wafer surface, or the wafer temperature, or both, during the programmed rate leg. In this work, the deposition rate is adjusted by changing the temperature, since this has a large effect on reaction rate.

The instantaneous differential step coverage Γ is defined by $\Gamma = \frac{dX(\tau)}{d\tau}/\frac{dL(\tau)}{d\tau}$. Because $L = R(p(Z,T),T(t))$, $\dot{X} = R(p(0,T),T(t))$ and $2X(t) + W(Z,t) = W_0$, see Figure 1, we obtain

$$\Gamma = \left[\frac{\theta(\xi_b,\tau)}{\theta(0,\tau)}\right]^{0.5} \qquad (14)$$

Since a path conforming to the PRCVD during the programmed rate leg of the process corresponds to constant Γ, equation (12) results in an expression for $\phi(\tau)$.

During the programmed rate leg of PRCVD, the temperature is determined from equations (12), (13) and (14), using a first order approximation described in Ref. [5]. During the constant rate leg (after SP), the temperature is held constant and equal to the temperature achieved at the end of the programmed rate leg. A reduction of processing time using PRCVD versus CRCVD depends on the selected

values of the process path parameters Γ and SP. For a given step coverage, however, Γ and SP may be found by trial and error so as to minimize processing time [3, 5]. One such method is to first fix Γ and then choose the value of SP which achieves the desired final step coverage. A minimum time is then obtained by searching over the range of Γ for the one yielding the minimum total time. Figure 2 shows one of the possible PRCVD paths for SiO$_2$ deposition by TEOS decomposition.

4 Optimally Controlled Chemical Vapor Deposition (OC-CVD)

Problem Formulation

To apply optimal control theory to the CVD process, we describe the process model as a set of ordinary differential equations. Since the step coverage is evaluated as the ratio of the thickness at the base of feature to the thickness at the wafer surface, we use the equations which describe the growth rate at the wafer surface and the growth rate at the base of the trench as the process model. For SiO$_2$ deposition by TEOS deposition, the simplified model of the process is written as

$$\frac{dL(t)}{dt} = R = 38.77 \exp\left(-\frac{E_a}{R_g T}\right) p^{0.5}(0, t) \qquad (15)$$

$$\frac{dX(t)}{dt} = \Gamma(t)\frac{dL(t)}{dt} \qquad (16)$$

where equation (16) comes from the definition of the instantaneous step coverage Γ.

We can obtain the expression for Γ by using the first order perturbation solution. The first order perturbation solution of the pseudo-steady-state model provides a reasonable estimate of concentration and deposition profiles for the range of $\dot{\sigma}$ values (reaction rates) which yield conformal depositions, since the concentration does not vary by more than a few percent from feature mouth to feature base [3, 5].

The first order perturbation solution to equation (12) for the dimensionless concentration profile in a trench can be written as

$$\theta(1, \tau) = \theta(0, \tau) - \frac{\sigma \mathcal{H}(\tau)(\alpha + 1)}{\mathcal{D}(0, \tau)H_0 \alpha_0} \qquad (17)$$

where $\alpha = \alpha(Z_b, t)$. Combining equations (13), (14) and (17), we obtain the following expression for Γ

$$\Gamma(t) = \left[1 - \frac{4}{T_0}\left(\frac{\pi m_i}{SK_b}\right)^{\frac{1}{2}} T^{\frac{1}{2}}(\alpha + 1)\varphi(\alpha)\frac{R(p(0, \cdot), T(t))}{C_0}\right]^N \qquad (18)$$

where

$$\sigma(\alpha) = \frac{18 + 16\alpha(Z, t) + 2\alpha^2(Z, t)}{18 + 7\alpha(Z, t)} \qquad (19)$$

The above equations reveal that Γ depends on the variables T and α, since $R(p(0, t), T(t))$ depends only on the temperature $T(t)$ when the partial pressure $p(0, t)$ is constant. Thus, the optimal control can be cast as follows:

For the system model equations (15) and (16), find an optimal control T (temperature), to minimize the processing time t_f subject to the constraints

$$X(0) = L(0) = 0 \qquad (20)$$

$$\frac{X(t_f)}{L(t_f)} \geq \text{a given step coverage} \qquad (21)$$

$$L(t_f) = \text{a given thickness} \qquad (22)$$

where Γ is given by equation (18).

Method

Since the system model is described by nonlinear ordinary differential equations, we do not attempt to solve the nonlinear optimal control problem analytically. Instead, we employ numerical techniques to determine the optimal control trajectories as the solution to a two-point boundary-value problem. One such general method is the so-called variation of extremals [6]. However, since in its standard form this method is not directly applicable to solve the optimal control for the CVD process, a suitable modification needs to be derived. In order to achieve this, we begin by employing optimal control theory to determine necessary conditions for a control trajectory to be optimal. In general, the problem is to find an admissible control U^* in the feasible set Ω, that causes the system

$$\dot{X}(t) = a(X(t), U(t), t) \qquad (23)$$

to follow an admissible trajectory $X^*(t)$ that minimizes the performance index

$$J(U) = h(X(t_f), t_f) + \int_{t_0}^{t_f} g(X(t), U(t), t)dt \qquad (24)$$

and satisfies boundary conditions $X(t_0) = X_0$ and $X(t_f) = X_f$, where t_0 is the specified initial time and t_f is the unknown final time.

By defining the Hamiltonian function

$$\mathcal{H}(X(t), U(t), P(t), t) = g(X(t), U(t), t) + P^T(t)[a(X(t), U(t), t)] \qquad (25)$$

the necessary conditions for optimality can be written as follows [6]

$$\dot{X}^*(t) = \frac{\partial \mathcal{H}}{\partial P}(X^*(t), U^*(t), P^*(t), t) \qquad (26)$$

$$\dot{P}^*(t) = -\frac{\partial \mathcal{H}}{\partial X}(X^*(t), U^*(t), P^*(t), t) \qquad (27)$$

$$\mathcal{H}(X^*(t), U^*(t), P^*(t), t) = \min_{U(t) \in \Omega} \mathcal{H}(X^*(t), U(t), P^*(t), t) \qquad (28)$$

for all $t \in [t_0, t_f]$, with boundary conditions

$$X^*(t_0) = X_o \qquad (29)$$

$$X^*(t_f) = X_f \qquad (30)$$

$$\mathcal{H}(X^*(t_f), U^*(t_F), P^*(t_f), t_f) + \frac{\partial h}{\partial t}(X^*(t_f), t_f) = 0 \qquad (31)$$

In our problem, we are interested in finding an optimal temperature trajectory to minimize the processing time. Hence, the performance index is the processing time while the control is the temperature T, i.e.,

$$J(U) = \int_{t_0}^{t_f} dt \qquad (32)$$

where $U = T$. Comparing with equation (38), we obtain the following equations $h(X(t_f), t_f) = 0$ and $g(X(t), U(t), t) = 1$ and the boundary conditions

$$X^*(t_0) = X_0 \qquad (33)$$

$$X^*(t_f) = X_f \qquad (34)$$

$$\mathcal{H}(X^*(t_f), U^*(t_f), P^*(t_f), t_f) = 0 \qquad (35)$$

Since the $P^*(t_f)$ is not specified in our case, we cannot directly apply the previously mentioned variation of extremals algorithm. In order to develop a suitable modification, we consider the observed final state $X(t_f)$ as the result of the choice of initial costates $P(t_0)$ and final time t_f. We then guess $P^{(0)}(t_0)$ and $t_f^{(0)}$, and update them so that the observed $X(t_f)$ converges to the desired value. Thus, we finally obtain the following equations for updating $P^{(i+1)}(t_0)$ and $t_f^{(i+1)}$.

$$X_f - X^{(i)}(t_f) = \frac{\partial X}{\partial P_0}(t_f^{(i)}, P_0^{(i)})(P_0^{(i+1)} - P_0^{(i)}) +$$

$$\frac{\partial X}{\partial t_f}(t_f^{(i)}, P_0^{(i)})(t_f^{(i+1)} - t_f^{(i)}) \qquad (36)$$

$$0 = \frac{\partial \mathcal{H}}{\partial X}[X_f - X(t_f^{(i)})] + \frac{\partial \mathcal{H}}{\partial P}\frac{\partial P}{\partial P_0}(t_f^{(i)}, P_0)(P_0^{(i+1)} - P_0^{(i)}) \qquad (37)$$

The iteration is terminated when $\|X_f - X(t_f^{(i)})\| \leq \gamma$ is satisfied.

It should be emphasized that the above procedure for computing the optimal trajectory does not account for the possibility that, during the iterations, the instantaneous step coverage Γ may become a complex number, especially near closure. Although a systematic way to prevent this situation is feasible (e.g., by introducing an additional constraint in the optimal control problem), for the sake of simplicity we adopted a different approach. We use a strategy similar to that used for PRCVD in that we only find the optimal control during the first leg so as to minimize the corresponding time, while for the second leg the processing time is fixed and the temperature is held constant.
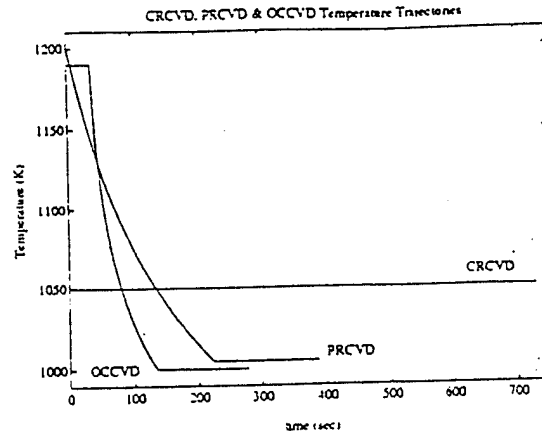

Figure 2: Temperature trajectory comparison under CRCVD, PRCVD and OCCVD

5 Simulation Results and Discussion

To compare the results of CRCVD, PRCVD and OCCVD, let us consider the model of SiO_2 deposition by TEOS decomposition, as described by Equations (15-16). The desired step coverage (SC) considered here is $SC = 0.96$. For best results in the PRCVD case, Reference [5] suggests the following choice of parameters: $\Gamma = 0.9775$ during the programmed rate leg and $SP = 0.899$. With these values and as shown in Figure 2, the temperature decreases until the switching point SP and remains constant thereafter. The total deposition time achieved using PRCVD, for 98 percent closure, is 386 seconds. For the same SC and percent closure, CRCVD requires 729 seconds of deposition time. The PRCVD process achieves 47% savings in processing time, compared with the CRCVD process which provides the same step coverage.

For OCCVD, we first select the time of the second leg as 100 seconds. This value is somewhat arbitrary and could be selected more systematically by performing an one-parameter optimization with respect to the second leg time. However, for the sake of simplicity, this approach is not adopted in the present study. With the temperature held constant in this leg, equal to the final temperature of the first leg in the previous iteration, equations (15–16) are integrated backwards in time to obtain the final state $X(t_f)$ at the end of the first leg. This final state is then substituted in equation (37) to yield the next estimate for $P(t_0)$ and t_f. These values are, in turn, used to integrate the state equations (26–27) forward in time with the input U (temperature) computed from equation (28). The procedure is repeated until convergence is achieved. The results of the OCCVD approach are shown in Figure 3. The first leg time is
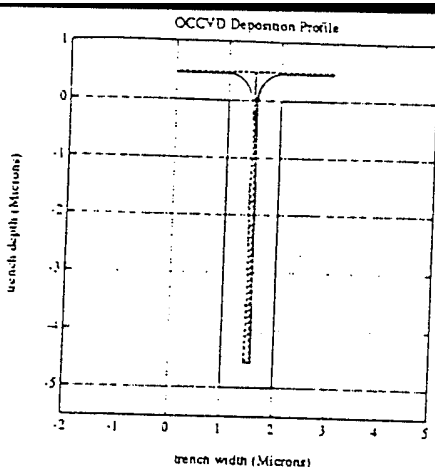
OCCVD Deposition Profile

Figure 3: Simulated final trench deposition profile under the OCCVD temperature trajectory. (PRCVD and CRCVD final profiles are similar). (a) solid: EVOLVE simulation. (b) dashed: simplified control model, used in the OCCVD computations.

178 seconds and the total time, for 98 percent closure, is 278 seconds. Compared to the PRCVD approach, OCCVD yields a 28% reduction of deposition time. The deposition time savings is 62% when compared with CRCVD (see Figure 2).

These estimated process time savings are validated by simulating the OCCVD, PRCVD and CRCVD process paths using the EVOLVE software [4]. The simulation results (Figure 3) show that the step coverage predicted using the simplified model differs from the more accurate EVOLVE predictions by 2.2% for OCCVD, 6.8% for PRCVD, and 4.6% for CRCVD. These results indicate that the time savings predicted using the simplified model remains qualitatively the same when a more accurate process model is used. The above results demonstrate that significant savings in processing time can be obtained, by using optimal control theory to compute temperature trajectories for a CVD process.

In practice, the time varying temperature predicted by OCCVD, will also affect partial pressures. As described in Ref. [7], reactor scale simulation could be used to determine the interaction between temperatures and partial pressures, and to establish reactor set-point trajectories. In contrast to the PRCVD approach, optimal control theory can easily account for the constraints imposed by the temperature dynamics in the solution (for example, to introduce constrains on the derivative of the temperature). These issues are left as topics of future work.

References

[1] M. K. Jain, *Maximization of Step Coverage at High Throughput During Low Pressure Deposition Processes.* Doctoral Thesis. Arizona State University, 1992.

[2] T. S. Cale. M. K. Jain and G. B. Raupp. "Programmed Rate Processing to Increase Throughput in LPCVD." *Journal of the Electrochemical Society,* Vol. 137, No. 5. may 1990.

[3] T. S. Cale, M. K. Jain and G. B. Raupp. "Tungsten and Other Advanced Metals for VLSI/ULSI Applications," *Materials Research Society,* Vol.V, Pittsburgh, PA. 1990. p.179.

[4] G. B. Raupp and T. S. Cale, *Chemistry of Materials,* 1, 207, 1989.

[5] J. H. Park. *Simulation of Low Pressure Chemical Vapor Deposition Using Combined Reactor Scale and Feature Scale Models.* Doctoral Thesis. Arizona State University, 1992.

[6] M. Knudsen, *The Kinetic Theory of Gases.* Wiley, New York, 1950.

[7] P. Clausing, *Physica.* 9. 65, 1929.

[8] A. C. Adams and C. D. Capio, *J. Electrochem. Soc.,* 126, 1042, 1979.

[9] D.E. Kirk. *Optimal Control Theory.* Prentice Hall, 1974.

[10] T.S. Cale. P.E. Crouch, L. Song and K.S. Tsakalis, *Proc. 1995 ACC,* Seattle, 1995

Optimal Control for Increasing Throughput in Low Pressure Chemical Vapor Deposition*

Lijuan Song[†], Sisan Shen[†], Kostas S. Tsakalis[†], Peter E. Crouch[†] and Timothy S. Cale[‡]
Arizona State University
[†] Center for System Science and Engineering
[‡]Center for Solid State Electronics Research
Box 877606, Tempe, AZ 85287-7606
c/o E-mail: tsakalis@enuxsa.eas.asu.edu

Abstract

In this paper we present the application of optimal control theory to the process of low pressure chemical vapor deposition (LPCVD) on patterned surfaces. An optimally controlled chemical vapor deposition (OC-CVD) protocol is developed by employing an approximate feature scale model, based on the simultaneous one-dimensional Knudsen diffusion and chemical reaction description of LPCVD, for the specific problem of maximizing throughput for a specified step coverage. The corresponding control conditions for the reactor are obtained by an adaptive solution of the inverse problem for the reactor scale model. Rigorous feature scale process simulations are used to verify that the trajectories of wafer temperature and reactant partial pressures provide good step coverages. Blanket Tungsten LPCVD is used to illustrate the methodology.

1   Introduction

In low pressure chemical vapor deposition (LPCVD) processes, compounds containing the species to be deposited react on the surface to form films. Single wafer Low Pressure LPCVD reactors are becoming more competitive, relative to multi-wafer LPCVD reactors, due to improvements in wafer-to-wafer deposition uniformity. A schematic diagram of a single wafer LPCVD reactor is illustrated in Fig. 1. To maintain device throughput in single wafer reactors at levels which are competitive with multiple wafer reactors, the deposition rate must be much higher. In LPCVD on patterned wafers, step coverage is a critical constraint. In general the step coverage deteriorates with increasing deposition rates. Thus, maintaining high step coverage while increasing the deposition rate is a big challenge in the design of LPCVD processes, and particularly in rapid thermal LPCVD. In conventional constant rate CVD (CRCVD), where the deposition conditions are held constant during the majority of the deposition process. Between the startup and shutdown transient, the deposition rate remains essentially constant at each location on the wafer surface during the process. It is well known that the step coverage decreases as feature aspect ratios increase, and the deposition rate during CRCVD must be low enough to ensure adequate step coverage. Although the final step coverage is an average of the step coverage obtained over the course of the deposition, it degrades most rapidly as feature close and the aspect ratios are highest [1]. Cale and coworkers have suggested programmed rate

CVD (PRCVD) for single wafer reactor LPCVD processes, in which the operating conditions are changed in a prescribed manner during the deposition [1, 2]. That is, the deposition rate decreases during processing, as feature aspect ratios increase. PRCVD can be used to decrease the deposition time for a given final step coverage, thereby increasing throughput, because the initial deposition rate is much higher than the CRCVD process which yields the same overall step coverage. Recent experimental work confirms the applicability of the PRCVD concept to blanket tungsten CVD using the hydrogen reduction of tungsten hexafluoride [3]. Films deposited with average rates which were over a factor of three higher than reference CRCVD processes had equivalent step coverages. In addition, all other PRCVD film properties were either as good as or better than the properties of CRCVD films.

In this paper, optimal control principles are employed, together with a combination of reactor scale and feature scale models, for the purpose of maximizing the process throughput, while maintaining a prescribed step coverage. Using a simplified feature scale model, optimal control inputs for the wafer are developed in an effort to minimize the processing time, subject to specified step coverage and final film thickness constraints. We use the processing temperature as the control (or manipulated) variable while keeping all other conditions constant for the feature scale model. Our simulation results show that the temperature trajectory obtained by solving the associated optimal control problem can yield a significant reduction of the processing time compared to the CRCVD approach, and some improvement relative to PRCVD processes.

In our feature scale analysis and simulations we assume that the species fluxes on the wafer surface remain constant. Such a situation does not occur naturally, since the conversion level of reactant species decrease as the deposition rate decrease. Thus, in order to achieve constant species fluxes, we employ an adaptive algorithm on the the reactor scale model to determine the reactor conditions for which, under the desired temperature trajectory, the species partial pressures remain constant. This procedure allows the derived optimal processing conditions on the wafer surface to be translated to optimal reactor-scale conditions that can be controlled through the available manipulated variables (temperature and flowrates). Future work will address issues related to relaxing the assumptions used for feature-scale simulation.

2 Feature-Scale Modeling of Single Wafer Tungsten LPCVD

A rigorous model of transport and reaction inside micron scale features during LPCVD is the "ballistic transport and reaction model" (BTRM) presented by Cale and coworkers [4, 5, 6, 7]. In the BTRM, three dimensional transport of species in free molecular flow and the reactions which consume or generate species are represented by Clausing-like integral equations. One accepted method to test, refine and validate models for transport and reaction kinetics, is to use a process simulator such as EVOLVE [8], which solves the BTRM equations, to simulate film deposition using these transport and reaction kinetic submodels. Simulated film profiles are compared with experimental film profiles; if the comparison is good, then it is likely that the models are satisfactory for engineering applications. The state-of-the-art procedure is to perform three dimensional (3-D) transport and reaction simulations on 3-D surfaces which can be represented in two dimensions by cross-sectioning (trenches and vias). These 3-D/2-D simulations can then be compared with SEM (or TEM) cross-section of trenches, lines, or vias with a high degree of rigor for CVD processes (though not without difficulties).

To develop our control model, we seek a model for transport and reaction in features during LPCVD which is easier to use than the BTRM. McConica and coworkers [9, 10] introduced the "diffusion and reaction model" (DRM) for free molecular flow and reaction in features. In the DRM, transport is modeled as one dimensional Knudsen diffusion in terms of local concentration gradients, and the deposition rate at each lateral position in the feature is determined by the concentration at that position. Cale and coworkers

extended the DRM [2, 11], and used it to analyze trends in conformality with changes in deposition conditions [1, 12] as well as to demonstrate the PRCVD concept [1, 2, 12]. Notice that the DRM is a "local" model, because the deposition at a point on the surface is governed by the concentrations in the vacuum next to that point, and the transport is governed by spatial derivatives. On the other hand, the BTRM is a "non-local" model, because depositing species can travel to any point from any other point in the feature or directly from the source volume, depending upon geometric visibility. It is clear that the BTRM is a better model, though the DRM is easier to manipulate and provides reasonable estimates for many low pressure CVD processes. Cale et al. [13] compared conformality predictions using simulators based upon the BTRM and the DRM, and found that they agree well as step coverage approaches unity. In this paper, we design our control scheme to maintain good conformality; that is, we keep gradients in concentration very small for the majority of the deposition. Thus, until very near closure, the DRM provides reasonable estimates of step coverage. After using the DRM to design our control strategy, we use simulations based upon the BTRM to validate the predicted protocol.

As in [1, 4, 12], the DRM used in this work employs a coordinate system that moves during deposition while the origin stays at the center of the feature mouth. Considering idealized symmetric features, such a coordinate system for a trench is shown in Fig.2, where

- $X(t)$ is the film thickness in the bottom of the trench at time t;

- $L(t)$ is the film thickness at the mouth of the trench at time t;

- $Z_b = H(t)$ is the instantaneous trench depth;

- $W(Z, t)$ is the instantaneous width of the trench at depth Z;

- W_0 is the initial trench width.

- H_0 is the initial trench depth.

Of course, such a deposition profile is a simplistic approximation which does not consume matter. However, it is a useful approximation to the results of more detailed simulations (see Fig. 3), and is adequate for the purpose of developing conditions for near-optimal processing using optimal control theory.

In our control model, we assume [1] that deposition occurs under conditions such that the rate depends only on the concentration of the limiting reactant, the feature is spatially isothermal, and that surface diffusion and radial or lateral gas concentration gradients are negligible. We treat the molecular transport as a one-dimensional process in which species flux is expressed in terms of local concentration gradients and Knudsen diffusion coefficients. The expressions for Knudsen diffusivity used in the DRM [2, 11] are based on cross-sectional averages for idealized feature geometry, e.g., infinitely long rectangular trenches or cylindrical contact holes. Here, we only consider the infinitely long rectangular trench model, although the same methodology is applicable to cylindrical contact holes as well [11]. Thus, the concentration of any species i at a depth Z in the feature is governed by

$$\frac{\partial}{\partial t}[A(Z,t)C_i(Z,t)] = \frac{\partial}{\partial Z}\left[A(Z,t)D_i(Z,t)\frac{\partial C_i(Z,t)}{\partial Z}\right] + P(Z,t)\sum_{j=1}^{r}v_{ji}R_j(Z,t) \tag{1}$$

where C_i is the instantaneous, local concentration of species i. A and P denote the feature cross sectional area available for molecular flow and feature perimeter at position Z and time t, respectively. v_{ji} is the generalized stoichiometric coefficient of species i in reaction j. $R_j(Z,t)$ is the rate of the j-th heterogeneous chemical reaction at position Z and time t. D_i is the instantaneous, cross sectional-averaged, local Knudsen

Dimensionless Variable	Definition
axial distance	$\xi = Z/H(t)$
time	$\tau = tD_{R0}/H_0^2$
feature depth	$\mathcal{H}(\tau) = H(t)/H(0)$
feature cross sectional area	$\mathcal{A}(\xi,\tau) = A(Z,t)/A(0,0)$
feature perimeter	$\mathcal{P}(\xi,\tau) = P(Z,t)/P(0,0)$
feature width	$\mathcal{W}(\xi,\tau) = W(Z,t)/W(0,0)$
Knudsen diffusivity	$\mathcal{D}(\xi,\tau) = D_i(Z,t)/D_i(0,0)$
concentration	$\theta_i(\xi,\tau) = C_i(Z,t)/C_i(0,0)$
rate of reaction	$G_j(\xi,\tau) = R_j(Z,t)/R_j(0,0)$
solid density	$\rho_{S_k} = C_{S_k}/C_{R0}$

Table 1: Definition of Dimensionless Variables

diffusivity of a gaseous species i. For infinitely long rectangular trenches, we use the following estimate of D_i [2]:

$$D_i(Z,t) = \left[\frac{8K_BT(t)}{\pi m_i}\right]^{1/2} \frac{H(t)}{4} \left[\frac{18 + 7\alpha(Z,t)}{18 + 16\alpha(Z,t) + 2\alpha^2(Z,t)}\right] \qquad (2)$$

where K_B is the Boltzmann constant, $T(t)$ denotes the temperature at the wafer surface at time t, m_i is the molecular mass of species i. The instantaneous, local aspect ratio α for rectangular trenches is

$$\alpha(Z,t) = \frac{H(t)}{W(Z,t)} \qquad (3)$$

The boundary conditions for the above second order partial differential equation are

$$C_i(0,t) = \frac{p_i(0,t)}{R_gT(t)} \qquad (4)$$

and

$$D_i(Z_b,t)\frac{\partial C_i(Z_b,t)}{\partial Z} = \sum_{j=1}^{r} v_{ji}R_j(Z_b,t) \qquad (5)$$

where R_g is the ideal gas constant, $p_i(Z,t)$ is the instantaneous partial pressure of species i at position Z and time t, and $Z=0$ is the mouth of the feature.

For the purpose of identifying the important parameters that dictate the step coverage and to determine their dependence on the CVD chemistry and operating conditions, the model equations are nondimensionalized [2]. The dimensionless variables are listed in Table 1. With these definitions, the model equations in dimensionless form are:

Material balance for species i

$$\frac{\partial}{\partial \tau}[\mathcal{A}(\xi,\tau)\theta_i(\xi,\tau)] = \frac{D_{io}}{D_{Ro}}\frac{1}{\mathcal{H}^2(\tau)}\frac{\partial}{\partial \xi}\left[\mathcal{A}(\xi,\tau)\mathcal{D}(\xi,\tau)\frac{\partial\theta_i(\xi,\tau)}{\partial \xi}\right] + \lambda_{Ri}\mathcal{P}(\xi,\tau)\sum_{j=1}^{r}v_{ji}G_j(\xi,\tau)\phi_j(\tau) \qquad (6)$$

Boundary conditions

$$\theta_i(0,\tau) = \frac{p_i(0,t)T(0)}{p_i(0,0)T(t)} = \frac{C_i(0,t)}{C_i(0,0)} \qquad (7)$$

$$\frac{\partial\theta_i(\xi_b,\tau)}{\partial \xi} = \frac{D_{Ro}C_{Ro}}{D_oC_{io}}\frac{A(0,0)}{H(0)P(0,0)}\frac{\mathcal{H}(\tau)}{\mathcal{D}(\xi_b,\tau)}\sum_{j=1}^{r}v_{ji}G_j(\xi_b,\tau)\phi_j(\tau) \qquad (8)$$

4

**Initial conditions**

$$\theta_i(\xi, 0) = 1 \tag{9}$$

$$\mathcal{H}(0) = 1 \tag{10}$$

where $0 \le \xi \le \xi_b$, $\tau > 0$, and $\xi_b = 1$. ϕ_j is referred to as the step coverage modulus for the reaction j and is given by

$$\phi_j(\tau) = \frac{2H^2(0)R_j(0,t)}{C_R(0,0)D_R(0,0)W(0,0)} \tag{11}$$

Finally λ_{Ri} denotes a partial pressure ratio and is defined as $\lambda_{Ri} = \frac{p_{Ro}}{p_{io}} = \frac{C_{Ro}}{C_{io}}$. The subscript "0" represents the position $\xi = 0$ and time $\tau = 0$, and the subscript "R" indicates the reference species.

For blanket tungsten LPCVD using the hydrogen reduction of tungsten hexafluoride, the reaction stoichiometry and rate expression are [14]:

$$3H_2 + WF_6 \longrightarrow W_2 + 6HF$$
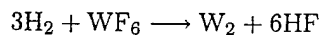
and

$$R(Z,t) = k_o \exp\left(-\frac{8300}{T(t)}\right) \frac{p_H^{1/2}(Z,t)p_F(Z,t)}{1 + k_p p_F(Z,t)}, \quad \left(\frac{mol}{cm^2 s}\right) \tag{12}$$

where p_F and p_H are the local partial pressures of tungsten hexafluoride and hydrogen in Torr, and T is the temperature in Kelvin. In the following, the subscript F represents tungsten hexafluoride and H represents hydrogen.

For this chemistry, the dimensionless species balances are [2]:

$$\frac{\partial \theta_F}{\partial \tau} = \frac{D_F(0,0)}{D_H(0,0)} \frac{1}{\mathcal{W}\mathcal{H}^2} \frac{\partial}{\partial \xi}\left(\mathcal{D}\mathcal{W}\frac{\partial \theta_F}{\partial \xi}\right) + \nu_F \lambda_{HF}\frac{\phi G}{\mathcal{W}} \tag{13}$$

and

$$\frac{\partial \theta_H}{\partial \tau} = \frac{1}{\mathcal{W}\mathcal{H}^2} \frac{\partial}{\partial \xi}\left(\mathcal{D}\mathcal{W}\frac{\partial \theta_H}{\partial \xi}\right) + \frac{\nu_H}{\mathcal{W}}\phi G \tag{14}$$

The boundary conditions are:

$$\theta_H(0,\tau) = \frac{p_H(0,t)T(0)}{p_H(0,0)T(t)} \tag{15}$$

$$\theta_F(0,\tau) = \frac{p_F(0,t)T(0)}{p_F(0,0)T(t)} \tag{16}$$

and

$$\frac{\partial \theta_H(\xi_b,\tau)}{\partial \xi} = \frac{3\mathcal{H}(\tau)}{2\alpha_o \mathcal{D}(\xi_b,\tau)}\phi(\tau)G(\xi_b,\tau) \tag{17}$$

$$\frac{\partial \theta_F(\xi_b,\tau)}{\partial \xi} = \frac{D_F(0,0)}{D_H(0,0)}\frac{p_H(0,0)}{p_F(0,0)}\frac{\mathcal{H}(\tau)}{2\alpha_o \mathcal{D}(\xi_b,\tau)}\phi(\tau)G(\xi_b,\tau) \tag{18}$$

The initial conditions are:

$$\mathcal{H}(0) = 1 \; ; \quad \theta_H(\xi,0) = 1 \; ; \quad \theta_F(\xi,0) = 1 \; ; \quad 0 \le \xi \le 1 \tag{19}$$

where G is the dimensionless rate of reaction, ϕ is the step coverage modulus given by equation (11) and α_0 is initial aspect ratio, given by $\alpha_0 = H_0/W_0$, ν_i is the generalized stoichiometric coefficient of species i ($\nu_F = -1$, $\nu_H = -3$), and λ_{HF} is

to follow an admissible trajectory $X^*(t)$, that minimizing a performance index of the form

$$J(U) = h(X(t_f), t_f) + \int_{t_0}^{t_f} g(X(t), U(t), t)dt \tag{38}$$

and satisfies the boundary conditions $X(t_0) = X_0$ and $X(t_f) = X_f$, where t_0 is the specified initial time and t_f is the unknown final time.

By defining the Hamiltonian function

$$\mathcal{H}(X(t), U(t), P(t), t) \triangleq g(X(t), U(t), t) + P^\top(t)[a(X(t), U(t), t)] \tag{39}$$

the necessary conditions for optimality can be written as [17]

$$\dot{X}^*(t) = \frac{\partial \mathcal{H}}{\partial P}(X^*(t), U^*(t), P^*(t), t) \tag{40}$$

$$\dot{P}^*(t) = -\frac{\partial \mathcal{H}}{\partial X}(X^*(t), U^*(t), P^*(t), t) \tag{41}$$

$$\mathcal{H}(X^*(t), U^*(t), P^*(t), t) = \min_{U(t) \in \Omega} \mathcal{H}(X^*(t), U(t), P^*(t), t) \tag{42}$$

for all $t \in [t_0, t_f]$, with boundary conditions

$$X^*(t_0) = X_o \tag{43}$$

$$X^*(t_f) = X_f \tag{44}$$

$$\mathcal{H}(X^*(t_f), U^*(t_f), P^*(t_f), t_f) + \frac{\partial h}{\partial t}(X^*(t_f), t_f) = 0 \tag{45}$$

If the costate $P(t_0)$ is known, equations (40)–(42) could be solved using numerical integration. Since this is not the case, we can guess values of $P^{(0)}(t_0)$ and $t_f^{(0)}$ for the initial costate and use them to numerically integrate (40)–(41) from t_0 to t_f. The observed values of $X(t_f)$ are then used to systematically adjust the guessed values of $P(t_0)$ and t_f. One technique for making systematic adjustments of the initial costate values and the final time is based on Newton's method for finding roots of nonlinear equations. Thus, the following equations are obtained for updating $P^{(i+1)}(t_0)$ and $t_f^{(i+1)}$.

$$X_f - [X(t_f)]_i = \left[\frac{\partial X(t_f)}{\partial P(t_0)}\right]_i [P^{(i+1)}(t_0) - P^{(i)}(t_0)] + \left[\frac{\partial X(t_f)}{\partial t_f}\right]_i (t_f^{(i+1)} - t_f^{(i)}) \tag{46}$$

$$0 = \left[\frac{\partial \mathcal{H}(t_f)}{\partial X(t_f)} + \frac{\partial}{\partial X(t_f)}\frac{\partial h(t_f)}{\partial t}\right]_i [X_f - [X(t_f)]_i] + \left[\frac{\partial \mathcal{H}(t_f)}{\partial P(t_f)}\frac{\partial P(t_f)}{\partial P(t_0)}\right]_i [P^{(i+1)}(t_0) - P^{(i)}(t_0)] \ldots$$
$$+ \left[\frac{\partial \mathcal{H}(t_f)}{\partial t_f} + \frac{\partial}{\partial t_f}\frac{\partial h(t_f)}{\partial t}\right]_i [t_f^{(i+1)} - t_f^{(i)}] \tag{47}$$

Thus, the update law is given by

$$\begin{bmatrix} P^{(i+1)}(t_0) - P^{(i)}(t_0) \\ t_f^{(i+1)} - t_f^{(i)} \end{bmatrix} = P_w^{-1} \begin{bmatrix} X_f - [X(t_f)]_i \\ -\left[\frac{\partial \mathcal{H}(t_f)}{\partial X(t_f)} + \frac{\partial}{\partial X(t_f)}\frac{\partial h(t_f)}{\partial t}\right]_i [X_f - [X(t_f)]_i] \end{bmatrix} \tag{48}$$

$$P_w = \begin{bmatrix} \mathcal{P}_X(P^{(i)}(t_0), t_f^{(i)}) & \left[\frac{\partial X(t_f)}{\partial t_f}\right]_i \\ \left[\frac{\partial \mathcal{H}(t_f)}{\partial P(t_f)}\right]_i \mathcal{P}_P(P^{(i)}(t_0), t_f^{(i)}) & \left[\frac{\partial \mathcal{H}(t_f)}{\partial t_f} + \frac{\partial}{\partial t_f}\frac{\partial h(t_f)}{\partial t}\right]_i \end{bmatrix} \tag{49}$$

8

The iteration is terminated when $\|X_f - X(t_f^{(i)})\| \leq \gamma$ is satisfied, where γ denotes the admissible error.

In the above equation, $\mathcal{P}_X(P^{(i)}(t_0), t)$ is the n by n matrix of partial derivatives of the components of $X(t)$ with respect to each of the components of $P(t_0)$, evaluated at $P^{(i)}(t_0)$; $\mathcal{P}_P(P^{(i)}(t_0), t)$ is the n by n matrix of partial derivatives of the components of $P(t)$ with respect to each of the components of $P(t_0)$, evaluated at $P^{(i)}(t_0)$, i.e.,

$$\mathcal{P}_X(P^{(i)}(t_0), t) \triangleq \left[\begin{array}{cccc} \frac{\partial x_1(t)}{\partial p_1(t_0)} & \frac{\partial x_1(t)}{\partial p_2(t_0)} & \cdots & \frac{\partial x_1(t)}{\partial p_n(t_0)} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial x_n(t)}{\partial p_1(t_0)} & \frac{\partial x_n(t)}{\partial p_2(t_0)} & \cdots & \frac{\partial x_n(t)}{\partial p_n(t_0)} \end{array} \right]_{P^{(i)}(t_0)} \tag{50}$$

$$\mathcal{P}_P(P^{(i)}(t_0), t) \triangleq \left[\begin{array}{cccc} \frac{\partial p_1(t)}{\partial p_1(t_0)} & \frac{\partial p_1(t)}{\partial p_2(t_0)} & \cdots & \frac{\partial p_1(t)}{\partial p_n(t_0)} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial p_n(t)}{\partial p_1(t_0)} & \frac{\partial p_n(t)}{\partial p_2(t_0)} & \cdots & \frac{\partial p_n(t)}{\partial p_n(t_0)} \end{array} \right]_{P^{(i)}(t_0)} \tag{51}$$

$\mathcal{P}_P$ is called the costate influence function matrix, and $\mathcal{P}_X$ is the state influence function matrix. Notice that equation (49) requires that $\mathcal{P}_P$ and $\mathcal{P}_X$ are known only at the terminal time t_f. The notation $[\cdot]_i$ means that the enclosed terms are evaluated on the i-th trajectory. Taking the partial derivatives of (40)-(41) with respect to the initial value of the costate vector and assuming that $\frac{\partial \dot{X}}{\partial P(t_0)}$ and $\frac{\partial \dot{P}}{\partial P(t_0)}$ are continuous with respect to $P(t_0)$ and t so that the order of differentiation can be interchanged, we obtain

$$\frac{d}{dt} \left[\mathcal{P}_X(P^{(i)}(t_0), t) \right] = \left[\frac{\partial^2 \mathcal{H}}{\partial P \partial X}(t) \right]_i \mathcal{P}_X(P^{(i)}(t_0), t) + \left[\frac{\partial^2 \mathcal{H}}{\partial P^2} \right]_i \mathcal{P}_P(P^{(i)}(t_0), t) \tag{52}$$

$$\frac{d}{dt} \left[\mathcal{P}_P(P^{(i)}(t_0), t) \right] = \left[\frac{\partial^2 \mathcal{H}}{\partial^2 X}(t) \right]_i \mathcal{P}_X(P^{(i)}(t_0), t) + \left[\frac{\partial^2 \mathcal{H}}{\partial X \partial P} \right]_i \mathcal{P}_P(P^{(i)}(t_0), t) \tag{53}$$

where $P^{(i)}$ obtained by integrating the reduced state-costate equations with initial conditions $X(t_0) = X_0$, $P(t_0) = P^{(i)}(t_0)$. The initial conditions for the influence function equations are

$$\mathcal{P}_X(P^{(i)}(t_0), t_0) = \left. \frac{\partial X(t_0)}{\partial P(t_0)} \right|_{P^{(i)}(t_0)} = 0 \tag{54}$$

$$\mathcal{P}_P(P^{(i)}(t_0), t_0) = \left. \frac{\partial P(t_0)}{\partial P(t_0)} \right|_{P^{(i)}(t_0)} = I \tag{55}$$

Equation (54) is derived by observing that a change in any of the components of $P(t_0)$ does not affect the value of $X(t_0)$, since the state values are specified at time t_0. Equation (55) follows from the observation that a change in the j-th component of $P(t_0)$ changes only $P_j(t_0)$.

Returning to our specific problem, we are interested in finding an optimal temperature trajectory which minimizes the processing time. Hence, the performance index is the processing time while the control is the temperature, T, i.e.,

$$J(T) = \int_{t_0}^{t_f} dt \tag{56}$$

Comparing with (38), we note that $h(X(t_f), t_f) = 0$ and $g(X(t), T(t), t) = 1$ and the boundary conditions are

$$X^*(t_0) = X_0 \tag{57}$$

$$X^*(t_f) = X_f \tag{58}$$

$$\mathcal{H}(X^*(t_f), T^*(t_f), P^*(t_f), t_f) = 0 \tag{59}$$

The update law for this problem is

$$\left[\begin{array}{c} P^{(i+1)}(t_0) - P^{(i)}(t_0) \\ t_f^{(i+1)} - t_f^{(i)} \end{array} \right] = P_w^{-1} \left[\begin{array}{c} X_f - [X(t_f)]_i \\ -\left[\frac{\partial \mathcal{H}(t_f)}{\partial X(t_f)}\right]_i [X_f - [X(t_f)]_i] \end{array} \right] \tag{60}$$

$$P_w = \left[\begin{array}{cc} \mathcal{P}_X(P^{(i)}(t_0), t_f^{(i)}) & \left[\frac{\partial X(t_f)}{\partial t_f}\right]_i \\ \left[\frac{\partial \mathcal{H}(t_f)}{\partial P(t_f)}\right]_i \mathcal{P}_P(P^{(i)}(t_0), t_f^{(i)}) & \left[\frac{\partial \mathcal{H}(t_f)}{\partial t_f}\right]_i \end{array} \right] \tag{61}$$

It should be emphasized that the above iterative procedure of computing the optimal trajectory involves solutions of nonlinear equations which may not be physically meaningful at every step. For example, especially during the iterations when the feature is near closure, a complex solution for the instantaneous step coverage Γ may occur. Although a systematic way to prevent this situation is feasible (e.g., by introducing an additional constraint in the optimal control problem), for the sake of simplicity we adopted a different approach. That is, we use a similar strategy as the PRCVD process [1, 2] in that we only find the optimal control during the first leg to minimize the corresponding time, while for the second leg the processing time is fixed and the temperature is held constant.

4 Reactor-Scale Modeling of Single Wafer Tungsten LPCVD

In this section, our objective is to determine the reactor-scale processing conditions, that yield the previously developed desired trajectories of deposition conditions at any point on the wafer surface. To achieve this, we consider a reactor-scale model with the manipulated variables (control inputs) being the susceptor temperature T (K), total pressure p_0 $(Torr)$ and flowrates of WF_6, H_2 and inert carrier gas $(sccs)$. Here, for simplicity, we assume that the control of the temperature is sufficiently good so that the prescribed temperature trajectory can be realized with adequate accuracy and spatial uniformity. The output variables (or controlled outputs) are the partial pressures of WF_6 and H_2. For control purposes, the modeling of such a process aims to establish relationships between the manipulated and output variables. Before any candidate models are proposed, experimental design and statistical analysis are usually helpful in determining a suitable model structure.

In the following we discuss the use of a simulation test bed rather than an actual reactor for model development. Simulation data are generated using a simulation platform called CFDSWR [18] [19], that simulates an LPCVD reactor. A full factorial experimental design is adopted and Yate's analysis [20] is applied to the simulation data in order to determine the most significant relationships between the manipulated variables and the output variables. In order to develop such a relationship, we observe that in a reaction-free environment the partial pressures p_F and p_H could be determined by the corresponding mole fractions and the total pressure. On the other hand, when chemical reactions occur, p_F and p_H at the wafer surface will decrease due to the consumption of the reactants during the reaction.

Assuming that the amount of reduction of each partial pressure is proportional to the apparent overall reaction rate for the corresponding species, we employ multiple response surface techniques to approximate these rates by an empirical expression of the manipulated variables. Thus, we arrive at the following empirical (but physically motivated) model, relating the manipulated variables with the reactant partial pressures at a point on the wafer surface

$$p_F(t) = q_{11} + q_{12}T(t) + q_{13}F_H(t) + q_{14}\frac{p_0(t)}{1 + F_H(t)} \tag{62}$$

10

$$p_H(t) = q_{21} + q_{22}T(t) + q_{23}F_H(t) + q_{24}\frac{p_0(t)F_H(t)}{1 + F_H(t)} \tag{63}$$

where q_{jk} are model parameters, T is the susceptor temperature, p_0 is the total pressure and F_H the flow rate of H_2. Notice that the "reaction feedback" modeling idea is invoked in selecting the parametric structure of the above model. That is, the last term in the above equations describes the reaction-free dependence of the partial pressures on the manipulated variables while the remaining terms describe the effect of the partial pressure reduction due to the reactions.

The model parameters q_{ij} are then estimated via a least squares algorithm to minimize the errors between the partial pressure trajectories of the reactants (WF_6 and H_2) as computed by the reactor-scale simulation platform CFDSWR [18, 19], and the trajectories predicted by the above empirical models. Of course, in practice, the parameters could be estimated using data obtained from experiments.

4.1 Recursive Computation of Flowrates in the Tungsten LPCVD Reactor

The OCCVD optimal temperature trajectory is obtained under the assumption that the partial pressures of WF_6 and H_2 on the wafer surface are kept constant. However, for constant flowrates and total pressure, the time variation of the optimal temperature trajectory induces a time variation in the partial pressures that has a detrimental effect on the deposition profile. In order to maintain constant partial pressures, we need to vary the other two reactor-scale manipulated variables, namely the total pressure p_0 and the H_2-flowrate F_H (the WF_6 flowrate is kept constant). To obtain the desired trajectories for which the partial pressures on the wafer surface remain constant under the OCCVD optimal temperature trajectory, we employ Newton's algorithm, based on our semi-empirical model. Such a problem is formulated by letting

$$y = f(u, \theta) \tag{64}$$

denote the simplified model (62)–(63), where $y = [p_F, p_H]^T$, $u = [p_0, F_H]^T$ and θ denotes the vector of model parameters, $\theta = [q_{11}, q_{12}, q_{13}, q_{21}, q_{22}, q_{23}]^T$. We assume that there exists a vector θ^* for which (64) is a "good" local approximation of the actual process, i.e., $y = f(u, \theta^*)$. Also, let y^* denote the desired trajectory to be achieved by varying u. Given an initial guess for u, we compute y using the simulation platform (CFDSWR). Then, u is iteratively updated by

$$u_{k+1} = u_k + \frac{\partial f^{-R}(u_k, \theta_k)}{\partial u}(y^* - y_k) \tag{65}$$

while θ_k is also updated in order to maintain the local fidelity of the approximation via an "indirect adaptation" of the form

$$\theta_{k+1} = \theta_k + \frac{\partial f^{-R}(u_k, \theta_k)}{\partial \theta}[y_k - f(u_k, \theta_k)] \tag{66}$$

until $|y^* - y_k| \le \epsilon$, for a given error threshold ϵ. In the above equations, the notation $(\cdot)^{-R}$ is used to signify the right-inverse of the corresponding matrix.

Finally, it should be mentioned that transients in the input and parameter updates could cause u to assume values outside the domain of validity and/or convergence of simulation model. As a remedy of this problem, projection techniques can be employed to limit the range of the inputs and the parameters ([21]). Notice that, in contrast to the input constraints, the parameter constraints may be difficult to determine. For this purpose, the approach of [22] can be used to obtain an initial estimate and update the parameter uncertainty set.

11

5 Results and Discussion

Here we consider tungsten deposition by hydrogen reduction of tungsten hexafluoride on patterned wafers, with $0.7 \times 3 \mu m$ trenches, modeled as described in previous sections. For a desired step coverage of $SC = 99.5\%$ and constant partial pressures on the wafer surface ($p_F = 0.125$ Torr, $p_H = 1.875$ Torr), under the optimal control CVD (OCCVD) approach we select the time of the second leg as 11 seconds. This value is somewhat arbitrary and could be selected more systematically by performing a one-parameter optimization with respect to the time of the second leg time. For the sake of simplicity, this approach is not adopted in the present study. With the temperature held constant in the second leg, equal to the final temperature of the first leg in the previous iteration, (40)-(41) are integrated backwards in time to obtain the final state $X(t_f)$ at the end of the first leg. This final state is then substituted in (60) to yield the next estimate for $P(t_0)$ and t_f. These values are, in turn, used to integrate the state equations (40)-(41) forward in time with the input U (temperature) computed from (42). The procedure is repeated until convergence is achieved. The results of the OCCVD approach are shown in Fig. 3. The first leg time is 14 seconds and the total time, for 98 percent closure, is 25 seconds. Under the same partial pressure conditions, the temperature trajectories for OCCVD and CRCVD approaches are shown in Fig. 4. Compared to the CRCVD approach, OCCVD yields a 34 percent reduction of processing time.

The above results demonstrate that significant savings in processing time can be obtained, without compromising the step coverage constraint, by using optimal control theory to compute temperature trajectories for a CVD process. The price paid for these improvements is the increased computational complexity of the solution and the difficulty in implementing the optimal trajectories on the actual process. Without attempting to completely resolve these issues here, we note that the former is of lesser significance, especially in view of the computational power of modern computers.

On the other hand, the implementation of the optimal trajectories may give rise to a more classical control problem, whose solution may be iterated together with the computation of the optimal trajectory. Consider, for example, the same CVD process and suppose that the wafer temperature is adjusted by a local temperature controller. The formulation of a similar optimal control problem for this case would use the desired wafer temperature (reference input or set-point in the local temperature controller) as the manipulated (control) variable. Assuming a simple (e.g., first or second order model) dynamic description of the relationship between the desired to actual wafer temperature, we could then proceed to find the optimal desired and actual temperature trajectories as before. Following this step, we can then design a temperature controller whose objective is to achieve the assumed dynamic description between the optimal reference trajectory and the actual wafer temperature, in the presence of modeling errors and/or disturbances. The essence of this description is to introduce a constraint on the speed of variation of the actual wafer temperature in the optimal control problem formulation. Another possibility would be to introduce constraints on the derivative of the temperature. The selection of the simple model should be such that its bandwidth (and, consequently, that of the actual wafer temperature) is within the bandwidth of the local temperature closed-loop. Thus, although the exact relationship between the desired and the actual wafer temperatures is very complicated or even unknown, the proper use of the local feedback controller can ensure the successful implementation of the optimal temperature trajectory.

Next, for the implementation of the OCCVD approach discussed here, the partial pressures on the wafer surface must be maintained constant, as assumed during the solution of the optimal control problem. Solving the corresponding inverse problem for the reactor scale model, we compute the desired trajectories for the manipulated variables p_0 and F_H that correspond to the desired partial pressures and optimal temperature trajectories (Fig. 5 and Fig. 6). As expected, the simulation of the reactor-scale model with

these manipulated variable trajectories yields nearly constant partial pressures on the wafer surface (Fig. 7 and 8). The resulting deposition profiles, simulated with EVOLVE, are shown in Fig. 9, for the uncontrolled case, and Fig. 10 for the controlled case. Notice that, although for this problem the differences in step coverage are relatively small (99.7% and 99.1% for the controlled and uncontrolled cases, respectively), a significant deterioration in the step coverage and trench closure may be observed in general, if the reactor-scale control is not implemented.

Finally, a more subtle problem arises from the need to establish such an optimal trajectory generation procedure as a practical alternative to the current approaches. Due to its complexity, the above procedure should be "automated" to a great extension in order to become usable by process engineers (e.g., provide the ability of easily generating a new optimal trajectory when the processing conditions change). Also notice that the proposed approach is "open loop" and requires relatively accurate models for its successful application. Although a quick remedy is to use on- or off-line adaptive techniques to ensure the accuracy of the model, an attractive alternative would be a closed loop design using other measurable quantities (e.g., deposition rates, outlet concentrations) from which the partial pressures of H_2 and WF_6 at the wafer can be inferred. These issues, however, extend beyond the scope of this study and are left as a topic of future research.

References

[1] T. S. Cale, M. K. Jain and G. B. Raupp, "Programmed Rate Processing to Increase Throughput in LPCVD," *Journal of the Electrochemical Society*, Vol. 137, No. 5, p.1526, 1990.

[2] M. K. Jain, *Maximization of Step Coverage at High Throughput During Low Pressure Deposition Processes.* Doctoral Thesis. Arizona State University, 1992.

[3] K. M. Tracy, "Programmed Rate Chemical Vapor Deposition: Blanket Tungsten Film Characterization," *M.S Thesis,* Arizona State Univ, May 1996.

[4] T. S. Cale and G. B. Raupp, "Unified Line of Sight Model for Transport and Deposition in Rectangular Trenches," *J. Vac. Sci. Technol.*, B 8(6), p.1242, 1990.

[5] T. S. Cale, T. H. Gandy and G. B. Raupp, "Free Molecular Transport and Deposition in Trenches," *J. Appl. Phys.*, 68(7), p.3645, 1990.

[6] T. S. Cale and G. B. Raupp, "Free Molecular Transport and Deposition in Cylindrical Features," *J. Vac. Sci. Technol.*, B 8(4), p.649, 1990.

[7] T. S. Cale, T. H. Gandy and G. B. Raupp, "A Fundamental Feature Scale Model for Low Pressure Deposition Processes," J. Vac. Sci. Technol., A 9(3), p.524, 1991.

[8] T. S. Cale, EVOLVE 4.0b, A low pressure deposition simulator, August 1994.

[9] C. M. McConica and S. Churchill, in "Tungsten and Other Refractory Metals for VLSI Application III," V. A. Wells, Editor, p.257, MRS Publishers, Pttsburgh, PA, 1988.

[10] C. M. McConica, S. Chatterjee and S. Sivaram, "Proc. 5th IEEE VLSI Multilevel Interconnection Conf." p.268, 1988.

[11] G. B. Raupp and T. S. Cale, "Step Coverage Prediction in Low Pressure Chemical Vapor Deposition," in *Chem. of Mat.*, 1, 207, 1989.

[12] T. S. Cale, M. K. Jain and G. B. Raupp, "Programmed Rate Processing to Increase Single Wafer Reactor Throughput in Blanket Tungsten LPCVD," in *Tungsten and Other Advanced Metals for VLSI/ULSI Applications*, eds., S. S. Wong and S. Furukawa, Materials Research Society, Vol.V, Pittsburgh. PA, 1990, p.179.

[13] T. S. Cale, G. B. Raupp, T. H. Gandy and M. K. Jain, "Comparison of Continuum Diffusion-Reaction and Fundamental Particle Flux Based Models for LPCVD," in *Tungsten and Other Advanced Metals for ULSI Applications in 1990*, eds., G. C. Smith and R. Blumenthal, Materials Research Society, p.165, 1991.

[14] J.H. Park and T.S. Cale, "Reactor Scale and Feature Scale Simulation of Programmed Rate CVD," *The 1st International Rapid Thermal Processing Conference*, p.393, 1993.

[15] M. K. Jain, T.S. Cale and T. H. Gandy, "Comparision of LPCVD film Conformalities Predicted by Ballistic Transport-Reaction and Continuum Diffusion-Reaction Models," *Journal of the Electrochemical Society*, Vol. 140, No. 1, p.242, 1993.

[16] A. Inamdar, and C. M. McConica, in *Tungsten and Other Refractory Metals for VLSI Applications IV*. R. A. Blewer and C. M. McConica, Editors, p.197, MRS Publishers, Pittsburgh, 1989.

[17] D.E. Kirk, *Optimal Control Theory*. Prentice Hall, 1974.

[18] J.H. Park, *User's guide for CFDSWR: Computational Fluid Dynamics of a Single Wafer Reactor*. Arizona State Univ. 1993.

[19] J. H. Park, *Simulation of Low Pressure Chemical Vapor Deposition Using Combined Reactor Scale and Feature Scale Models*. Doctoral Thesis. Arizona State University, 1992.

[20] R. E. Walpole, and R. H. Mayers, *Probability and statistics for Engineers and Scientists*. Macmillan Publishing Company, 1985.

[21] S. Sastry, and M. Bodson, *Adaptive Control Stability, Convergence and Robustness*. Prentice Hall. 1989.

[22] K.S. Tsakalis and L. Song, "Set-membership estimation for weakly nonlinear models: An application to the adaptive control of semiconductor manufacturing process," *33rd CDC*, 1994.
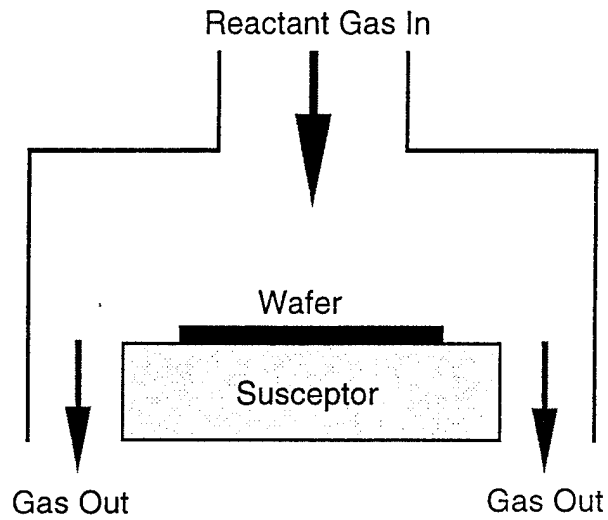
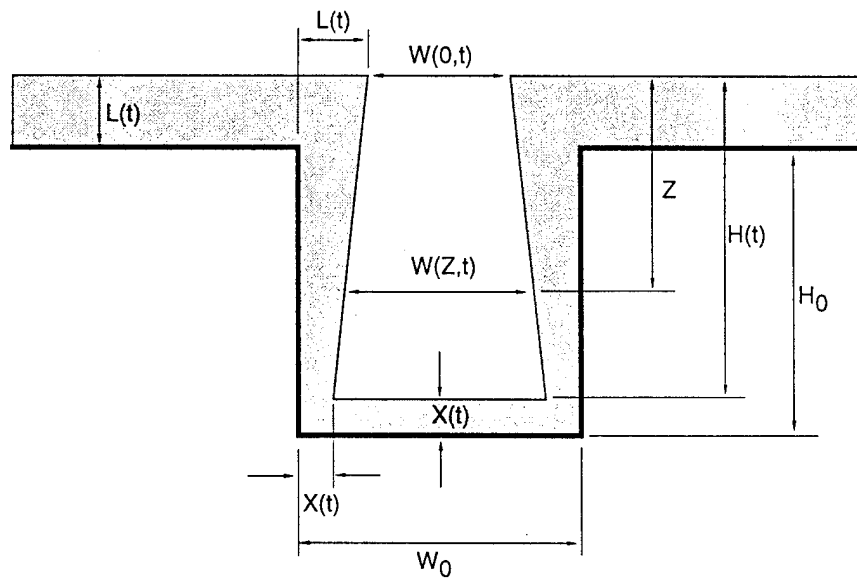Figure 1: Schematic diagram of the Single Wafer LPCVD reactor.

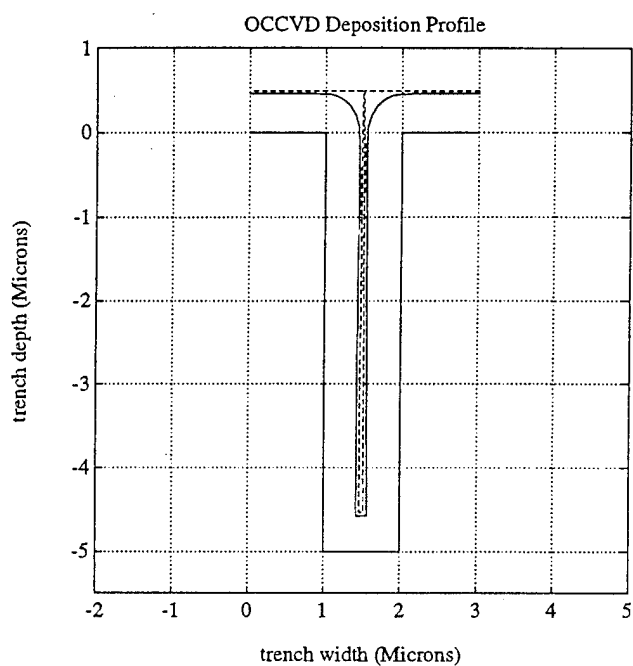Figure 2: Idealized cross section of a feature during deposition.

16

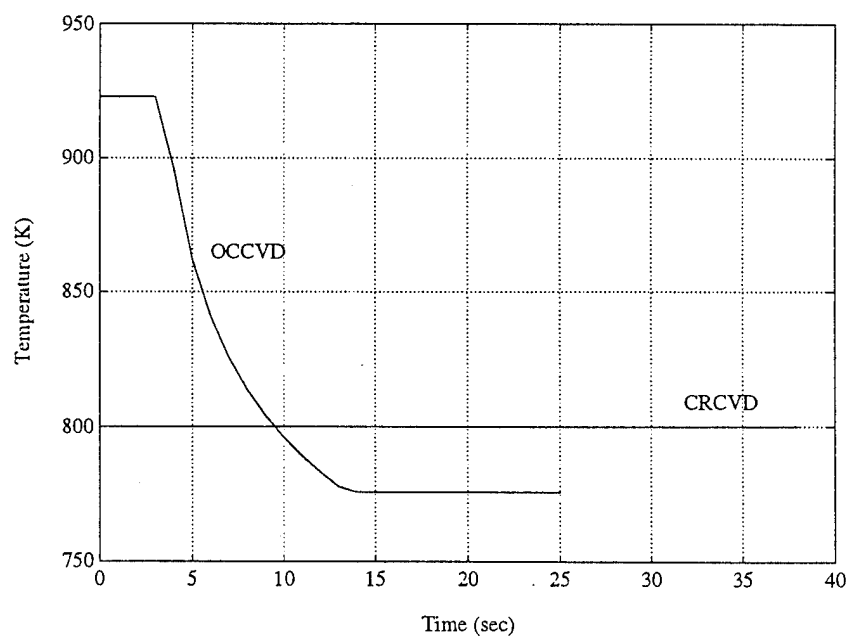Figure 3: Deposition profiles under OCCVD (solid: full feature-scale model (EVOLVE); dashed: simplified model).

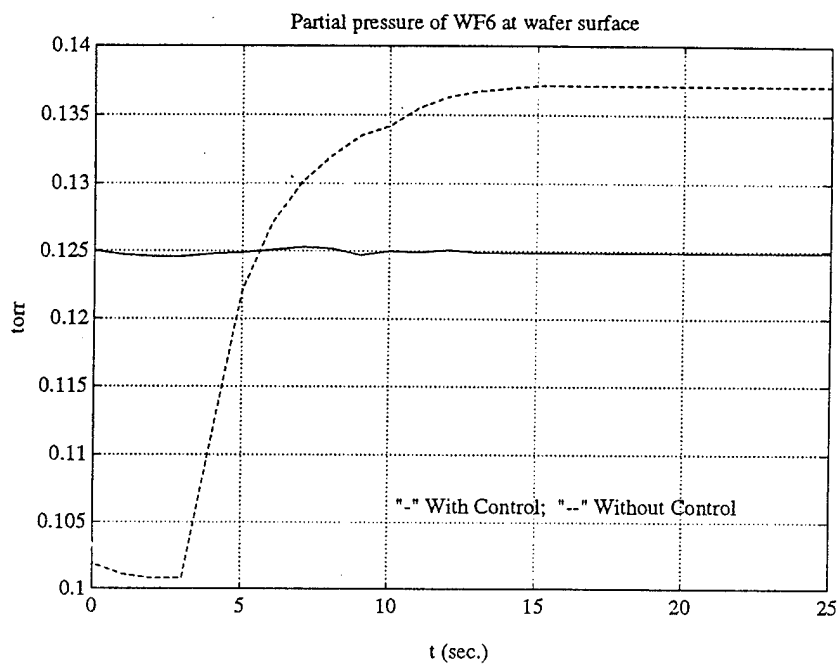Figure 4: Temperature trajectory comparison under CRCVD and OCCVD.

16

Figure 7: Comparison of the resulting partial pressure trajectories for WF_6 at the wafer surface (constant versus controlled total pressure and H_2 flowrate).
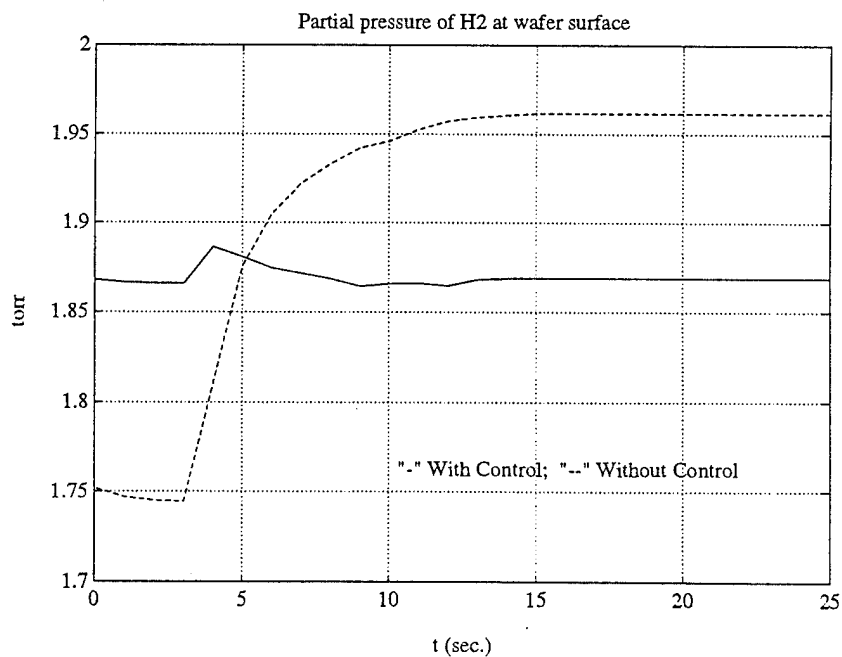

Figure 8: Comparison of the resulting partial pressure trajectories for H_2 at the wafer surface (constant versus controlled total pressure and H_2 flowrate).
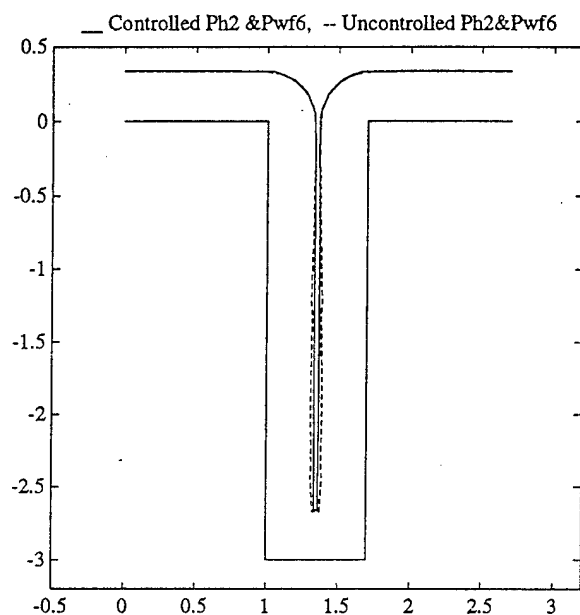
18

Figure 9: Comparison of resulting deposition profiles (EVOLVE simulation): Constant versus Controlled Reactor variables (Total Pressure and H_2 Inlet Flowrate)
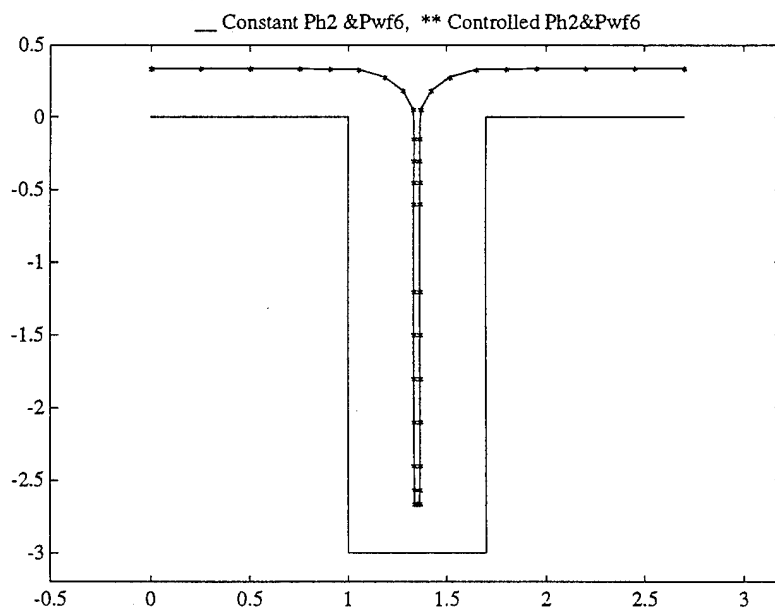

Figure 10: Comparison of resulting deposition profiles (EVOLVE simulation): Constant partial pressures versus resulting partial pressures (Controlled by Reactor variables Total Pressure and H_2 Inlet Flowrate)